DOCTOR OF MEDICINE

The Effect of an Integrated Model of Dialogic Feedback with Encouraged Self-Regulation on Psychomotor Task Performance and Learner Experience and Understanding of Feedback

Gill, Sarah

Award date:
2019

THE EFFECT OF AN INTEGRATED MODEL OF DIALOGIC FEEDBACK WITH ENCOURAGED SELF-REGULATION ON PSYCHOMOTOR TASK PERFORMANCE AND LEARNER EXPERIENCE AND UNDERSTANDING OF FEEDBACK.

MISS SARAH L GILL

MD in Medical Education

University of Dundee

May 2019

TABLE OF CONTENTS

PAGE

ACKNOWLEDGEMENTS

I would like to thank Fraser Harrold, Sean McAleer and Rola Ajjawi for their supervision and mentorship during this research degree. Their patience, understanding and considerable skill made this thesis possible. I was very fortunate to have great Medical Education colleagues, Katy Orr and Richard Steven, during my fellowship; I benefitted so many times from their insight and straight-forward advice. I would also like to thank my Orthopaedic colleagues for their support during a time of intense workload. Finally, I would like to thank my family for keeping their faith!

DECLARATION

I have consulted all references cited; undertaken the work of which the thesis is a record; and this thesis has not been previously accepted for a higher degree.

………………………………………………………………………………………………..

Miss Sarah L Gill

SUMMARY OF THESIS

In the first chapter of this thesis, the topic of feedback is framed. The importance of feedback within medical education is asserted and the concept of a reconceptualisation within wider education is introduced.

The second and third chapters represent highly structured literature reviews of feedback and self-regulation respectively within medical education. They illustrate the preoccupation of establishing the quantitative effect of feedback via experimental studies and significant limitations of the literature available. They highlight the relative paucity of information available in relation to the effect of self-regulation on learning. Finally, they introduce the emerging reconceptualisation of feedback within medical education but the limited understanding of how to introduce a dialogic feedback model and its unknown effects.

Chapter four presents the concept of an integrated model of dialogic feedback with encouraged self-regulation, presents the research questions chosen for this study, and discusses important over-arching design considerations for the presented research.

Chapters five and six represent the design, execution and analysis of two pilot studies. These pilot studies offered practical experience, were instrumental in the maturation of researcher understanding of the subject matter and enabled robust statistical design of the final study. The research questions relating to the final study are contained the end of chapter six.

Chapter seven describes in detail the final study design and methods, including methods of quantitative analysis.

Chapter eight details the quantitative results of the final study. It describes the success of the randomised control design in limiting bias and details the statistical analysis. The quantitative analysis illustrates the improved intra-visit and cross-over task performance associated with the dialogic feedback model.

Chapter nine explores the thematic analysis of the learner experience and perceptions at the end of the study. It provides evidence that engagement in a dialogic feedback model promotes an active learner role, cognitive engagement, and increasing perceptions of self-efficacy.

Discussion and conclusions are presented in chapter ten. It presents the key study findings, in addition to appraisal of the study and identification of important related future research.

CHAPTER 1: INTRODUCTION AND THE RECONCEPTUALISATION OF FEEDBACK

1.1 A BRIEF INTRODUCTION TO THE RESEARCHER

I am a current Trauma and Orthopaedic surgery trainee in the East of Scotland. I have received the entirety of my education in the UK, graduating from Dundee Medical School in 2007. As a junior doctor, I worked in Edinburgh on clinical rotations through various primary and secondary care departments before commencing my Orthopaedic training in 2009. Between February 2014 – 2016, I undertook a Fellowship in Medical Education at Dundee Medical School and an MD in Medical Education.

This research opportunity allowed me to consider a specific area of medical education. Feedback is of great importance in surgical training. It occurs in formal and informal episodes but there is little evidence of medical education theory playing a major role in how these episodes form part of the training curriculum. Therefore, interest in the educational theories surrounding feedback was the starting point for this research project. As a trainee, I wondered how research in this area might influence future surgical training; as a surgeon, I wanted to know the quantitative effects of feedback on learning; and later as a maturing researcher in medical education, I became interested in the relationship between the different models of feedback and the learner's experience and understanding of feedback.

1.2 THE IMPORTANCE OF FEEDBACK

Feedback is widely accepted as a very important contributor to the process of learning (Carless, 2006). As learners, we seek to obtain it; as tutors, we seek to provide it. Educational interest in this potentially powerful tool has increased over the last decade. In the PubMed electronic database, there were 2332 references matching the MESH terms 'feedback' and 'education' between July 2004 – 2014 and 375 (16%) were in the last 12 months of that period. Similarly, feedback as a subject is of increasing interest in medical education.

Indeed, the effectiveness of feedback cannot be overlooked. Combined meta-analyses exploring the influence of over 100 different factors on student educational achievement (Hattie & Timperley, 2007) concluded that the average effect size of feedback (0.74) given in

the classroom was twice the average size of factors relating to learning, such as students' prior cognitive ability, socioeconomic influences, the use of homework and class size.
Meta-analyses exploring the effect of feedback on clinical performance (Veloski et al, 2006) found that 74% of studies associated feedback with a positive impact. Deeper analysis reported on the effect of different characteristics of feedback and suggested that source and duration of feedback were the most relevant to its influence, overshadowing other variables such as individual versus group feedback, dependency level of the recipient and the degree of privacy of disclosure.

Interest in feedback within medical education and clinical medicine is sustained by the apparent inaccuracy of self-assessment and appraisal. There is a poor correlation between trainee self-assessment and assessment from external sources in the clinical environment in terms of the assigned level of performance (Lipsett et al, 2011). High-performing trainees tend to underestimate skill level and poor-performers conversely tend to over-estimate it. Therefore, identifying learning needs and basing development on self-assessed competency would appear inaccurate and of limited benefit without an external source of information to compare against.

However, exploration of the medical and surgical education literature pertaining to feedback, as detailed in following sections of this thesis, reveals that the overall quality of existing quantitative experimental studies is poor. This limits what the existing body of literature can tell us about the effect of traditional information-transmission feedback on learner performance.

1.3 FEEDBACK IN EDUCATION

In recent years, the case has been made for a reconceptualisation of the feedback process within education. Carless et al (2011) explored a model of dialogic feedback. Their premise was that feedback in education must move away from a process of one-way information transmission to the creation of feedback dialogue between these parties. In this work, Carless et al draw together several themes. They proposed that the student role within the feedback process must be enhanced and the onus moves away from the teacher's role in delivering feedback, towards the learner's role in viewing and digesting separate episodes of feedback to create a more holistic picture. In this way, they extend the role of feedback beyond only to the exact context in which it is given and offers the notion of 'exploratory' rather than 'directive' feedback.

'Feedback…. is not just the feedback you give to students in written form. Feedback is a kind of support which gives students a sense of scaffolding and will gradually get them to be more independent.' (p. 402)

This model supports the principle championed by Riordan and Loacker (2009), that through active engagement in self-assessment, students will *'become independent lifelong learners who have learned from us but no longer depend on us to learn'* (p. 181).

This evolution of the feedback process does not remove the external input of the teacher but positions the student centrally within the learning and feedback process. Preservation of the role of trainer input to correct trainee errors is supported by the literature. Trainers are able to more rapidly address 'faulty perceptions' adopted by trainees compared to the trainee themselves (Hattie & Timperley, 2007) and are quicker to spot trainee errors (Nicol, 2007).

Boud and Molloy (2013) further championed the learner's role. In this landmark paper, the authors define two models of feedback: mark one and mark two feedback. Mark one is akin to the *'paradigm of telling'*, with the focus on the external provision of performance-based information. In contrast, mark two feedback requires *'the active positioning of learners as elicitors of knowledge for improvement, not just the recipients of input from others.'* (p. 705) This idea is an extension of Nicol's previous work (2009), in which it was suggested that for it to be beneficial, students must use feedback from teachers as a base for self-assessment. That is, *'they must decode the message, internalize it and use it to make judgments about and modify their own work'* (p. 207).

The promotion of the learner from a passive to an active role may be achieved through the adoption of self-regulation. Zimmerman (2000) defined self-regulation as *'self-generated thoughts, feelings and actions that are planned and cyclically adapted to the attainment of personal goals'* (p. 13). Self-regulation involves several key processes: forethought (with utilisation of process goals); performance (with active self-monitoring); and reflection (consideration of process to inform ongoing self-regulation cycles). There is evidence to support that goal setting positively impacts motor skill development and that self-monitoring positively influences self-efficacy beliefs (Zimmerman, 1996).

Therefore, combination of dialogic feedback with promotion of self-regulation advances feedback from a linear process of information transfer, from learned tutor to naïve learner, to an ongoing cycle of tutor- facilitated development driven by the learner themselves. It makes the active role of both parties in the process explicit. Tutor feedback is tailored towards supporting context-specific self-regulatory learner performance, and the learner uses this external information to aid internal appraisal and the formation of future performance modifications.

Despite this well-articulated paradigm shift in the wider education literature, review of the literature as presented in this thesis, illustrates that a dialogic model of feedback is currently only in the concept stages within medical and surgical education. The surgical education literature remains preoccupied with exploration of the effect of tutor-generated and delivered feedback via experimentally designed studies. Within these studies, there is no exploration of the learner experience or role and no studies involved a dialogic feedback process. A slightly more advanced discussion within the medical education literature of new ideas and concepts have theorised a move towards a shared tutor:learner feedback dialogue (Carless et al, 2011; Boud & Molloy, 2013) but this is limited by the tutor-centric language that persists, and a lack of understanding of how this feedback is created (Rudland et al, 2013; Telio et al, 2015).

CHAPTER 2: STRUCTURED LITERATURE REVIEW OF FEEDBACK

The current chapter presents the findings of a structured systematic review of feedback literature in medical and surgical education. The focused aim of this literature review was to identify and analyse research articles published specifically to investigate the measurable effect of feedback, via an experimental design, and those that theorise feedback. In doing so, I intended to explore the conclusions offered in relation to the effect of feedback within medical and surgical education, present and critique the quality of the quantifiable evidence these were based upon, and to capture the existing discourse in this literature surrounding models of feedback. This review was essential in identifying important themes relating to feedback but also gaps in the literature that might represent areas which new research might usefully address.

2.1 FEEDBACK IN MEDICAL AND SURGICAL EDUCATION

A structured review was conducted to identify literature relevant to the study of feedback in medical and surgical education. The search was conducted via the PubMed database, a service provided by the National Centre for Biotechnology Information (NCBI) of the US National Library of Medicine. PubMed provides free access to MEDLINE®, the NLM® database of indexed citations and abstracts to medical, nursing, dental, veterinary, health care, and preclinical sciences journal articles. The PubMed journal list includes approximately 30,000 journals. The review was conducted in July 2016.

2.2 SEARCH TERMS AND RESULTS

Two separate searches using broad Medical Subject Headings (MeSH) terms were performed. This maximised capture of relevant articles, and the results were then combined:

- Search one: medical education AND feedback
- Search two: surgical education AND feedback

Box 1: Literature search criteria

Title review: Exclusion criteria	**Abstract review: Exclusion criteria**
No mention of feedback in the title Non-education-based feedback - Clinical handover/debrief - Patient feedback Feedback on educational courses Multisource feedback Video-based feedback Peer-to-peer feedbacks	Criteria pertaining to title reviews Abstract (or subsequent article) unavailable in English Abstract-only publications Focus on measuring prevalence of feedback Focus on importance of feedback Focus on perceptions of feedback Focus on elements affecting feedback-seeking Behaviour Focus on assessment including self-assessment, not feedback Focus on work-place based assessments Research based on non-interventional studies Focus on reporting of practice without investigation of intervention nor discussion of theory Study with no evaluation of feedback intervention Focus on development of feedback tool

The searches were purposefully broad to maximise the potential inclusion of articles of interest in subsequent stages of the literature review. Research article titles and then abstracts were reviewed and subjected to the inclusion and exclusion criteria detailed in Box 1. Duplicate articles were removed. Figures 1-4 illustrate the stepwise process through which the summative collection of literature was attained.

Figure 1: 'Medical education' AND 'feedback' systematic literature review

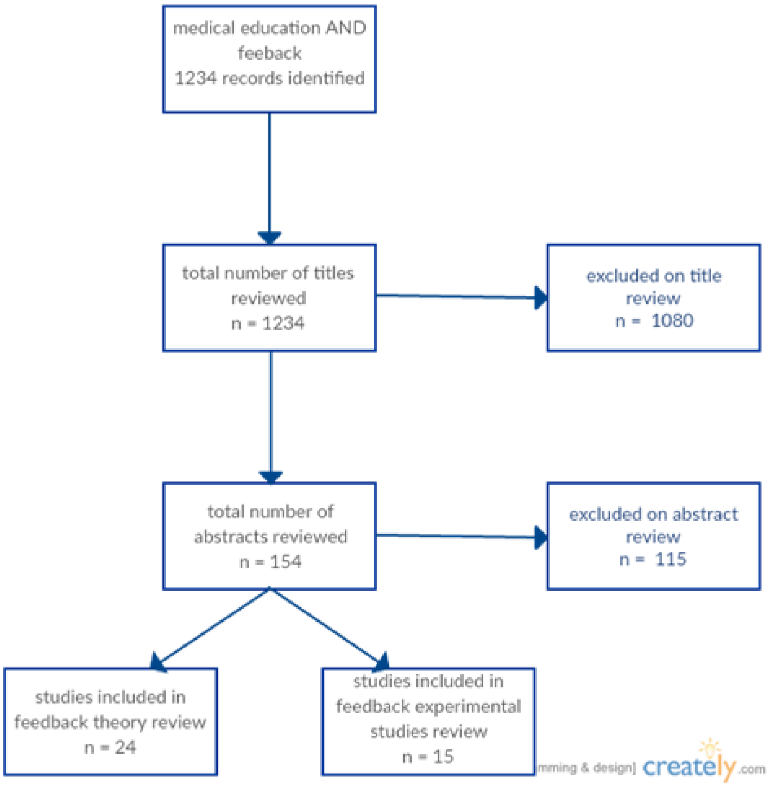


Figure 2: 'Surgical education' AND 'feedback' systematic literature review

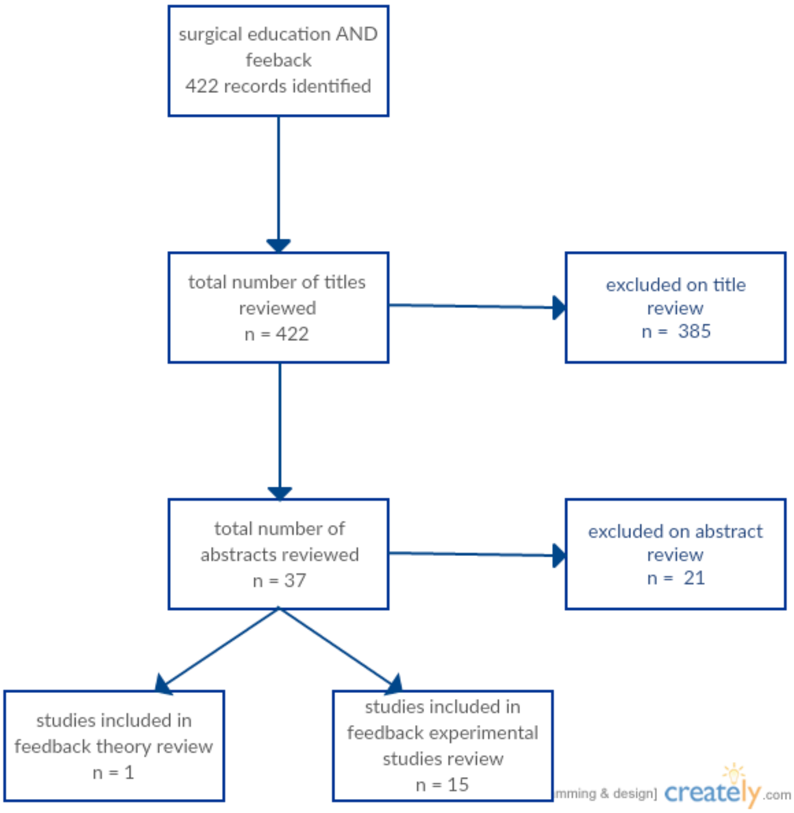


Figure 3: Feedback experimental studies systematic literature review

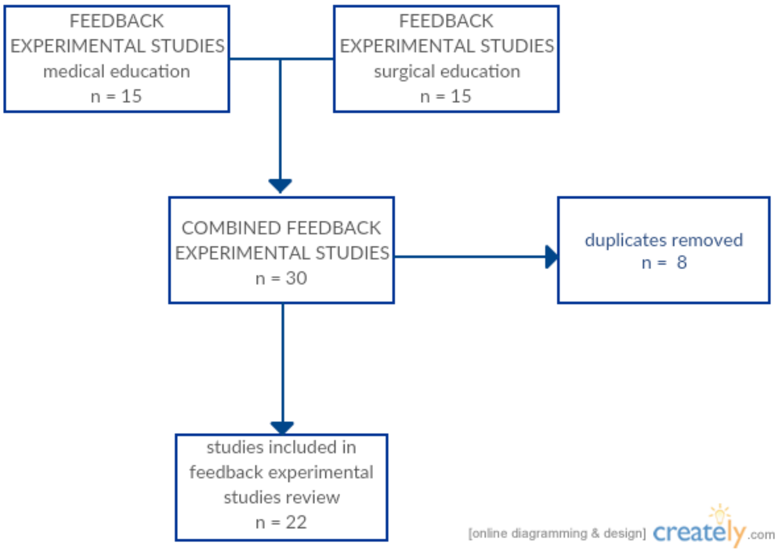
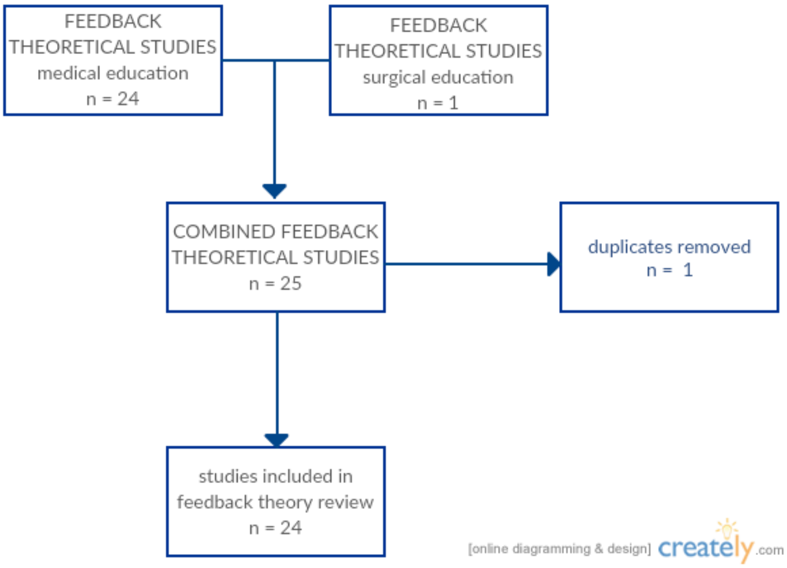


Figure 4: Feedback theoretical studies systematic literature review



2.3 THE INVESTIGATION OF FEEDBACK VIA EXPERIMENTAL STUDIES WITHIN THE MEDICAL AND SURGICAL EDUCATION LITERATURE

Via this systematic literature review, 22 papers were identified in which an interventional, experimental study design was used to investigate the effect of feedback within medical and surgical education.

There are 21 different first name authors of the 22 papers. The primary role of 16 authors is as a clinician, two are medical educationalists and three are non-medical education academics. The majority of these papers (n=13) were published in clinical journals; eight were published in medical education journals and one in an academic, non-education journal. This analysis identifies that it is largely clinicians, publishing for the benefit of peers via clinical journals that focus on the objective measurement of the effect of feedback. A summary of the 22 papers is found below in Table 1.

Table 1: Summary of interventional, experimental papers

Aim	Participants	Task	Intervention	Task measures	Design	Results
Rogers et al (2000) The impact of external feedback on computer-assisted learning for surgical technical skill training. Am J Surg, 179(4), 341-3.						
To explore the effect of external feedback on psychomotor task	*Undergraduate Novice* 105 medical students	*Psychomotor Bench model* Two-handed square surgical knot	*G1:* CAL resources only *G2:* Expert feedback during 1-hour computer assisted learning education session	*Efficiency:* Total time *Quality:* Proportion of square knots tied Investigator global rating score	*Single visit RCT* Pre-intervention performance 1-hour education + intervention Post-intervention performance	*Efficiency:* Both groups improved; no inter-group difference. *Quality:* no inter-group difference in number of square knots; G2 had higher scores in post-intervention global rating scores
Backstein et al (2004) The effectiveness of video feedback in the acquisition of orthopedic technical skills. Am J Surg, 187(3), 427-32.						
To explore the effect of no feedback, video-feedback (viewing) or expert-feedback	*Postgraduate* 29 Residents (orthopaedics)	*Psychomotor Bench models* Sawbone plating, TBW to olecranon, Z-plasty (porcine model)	*G1:* No feedback *G2:* Participant video review *G3:* Video review and expert feedback	*Quality:* Investigator global rating score Investigator task-specific score via checklist	*Single visit RCT* Task demonstrated by expert Pre-intervention performance Intervention Post-intervention performance Cross-over design; each participant was in each feedback group for one of the tasks	*Quality:* No inter-group differences in relation investigator global rating score nor investigator task-specific score via checklist
Boehler et al (2006) An investigation of medical student reactions to feedback: a randomised controlled trial. Med Educ, 40(8), 746-9.						
To evaluate learning outcomes and perception for feedback versus compliments	*Undergraduate* 33 medical students	*Psychomotor Bench model* Two-handed square surgical knot	*G1:* Generic scripted general compliments *G2:* Specific, constructive investigator feedback	*Quality:* Blinded expert task-specific score via checklist via video review Student satisfaction via Likert scale	*Single visit RCT* Pre-test performance Task instruction Pre-intervention performance Task practice + intervention Post-intervention performance Students rated satisfaction	*Quality:* No inter-group difference at pre-test or pre-intervention performance. *G2* significantly better at post-intervention performance *G1* significantly more satisfied with instruction

Aim	Participants	Task	Intervention	Task measures	Design	Results
Judkins et al (2006) Real-time augmented feedback benefits robotic laparoscopic training. Stud Health Technol Inform, 119, 243-8.						
To investigate the use of real time augmented feedback in virtual reality laparoscopic surgical task performance	*Undergraduate* 12 medical students	*Psychomotor Virtual reality simulated laparoscopic tasks:* Bimanual carry Needle passing Suture tying	*G1:* No feedback *G2:* Real-time on-screen grip strength indicator *G3:* Real-time on-screen instrument speed indicator *G4:* Real-time on-screen instrument coordination indicator	*Efficiency:* Time to complete Distance instruments travelled Mean speed of instruments *Quality:* Grip force	*Single visit RCT* Verbal task instructions 3 pre-intervention performances 10 training trails with intervention 3 post-intervention performances Repeated for each of the 3 tasks	All groups quicker in all tasks post vs pre-performance *G2* applied less force post- vs pre-performance in all 3 tasks. Complex interactions: when one factor is emphasised, others are neglected
Xeroulis et al (2007) Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. Surgery, 141(4), 442-9.						
To investigate the effect of computer-assisted learning and summary and concurrent feedback on learning	*Undergraduate* 60 medical students	*Psychomotor Bench model* Suturing (hand and instrument knot tying)	*G1:* No feedback *G2:* Self-directed study with demo video available (on average, participants viewed 3 times) *G3:* Concurrent (intra-performance) expert feedback *G4:* Summary (post-performance) expert feedback	*Efficiency:* Total time. Number of hand movements made. *Quality:* Global rating score via blinded expert video review	*Two visit RCT* Instructional video Pre-intervention performance 1-hour practice session (19 task trials) with last trial treated as immediate post-intervention performance. *Delayed retention performance 1-month post intervention.*	*Efficiency:* All groups improved at immediate post-intervention performance; *G2, G3* and *G4* > *G1*. Only *G2* and *G4* retained improvement at delayed testing. *Quality:* All groups improved at immediate post-intervention performance; *G2, G3* and *G4* > *G1*. Only *G2* and *G4* retained improvement at delayed testing.
Van Sickle et al (2007) The effect of escalating feedback on the acquisition of psychomotor skills for laparoscopy. Surg Endosc, 21(2), 220-4.						
To investigate the effect of type and quality of feedback on laparoscopic task performance	*Undergraduate Novice* 32 medical or science students	*Psychomotor* Practice: Maze-tracking task Test: Laparoscopic box trainer cutting task	Types of error reinforcement *G1:* No feedback *G2:* Buzzer when edges touched *G3:* Examiner says 'error' when edges touched *G4:* Both buzzer and voice when edges touched	*Efficiency & Quality:* Number of incorrect incisions made (errors) and number of correct incisions made	*Single visit RCT* 10 x 2-minute training trials (maze task) 1 x - minute laparoscopic cutting task performance	*Efficiency & Quality:* All feedback groups (*G2-4*) improved in the trial period All feedback groups (*G2-4*) superior in cutting task *G4* performed the best (both re reduced error and volume of correct incisions) No difference between *G2 & G3*

Aim	Participants	Task	Intervention	Task measures	Design	Results
Grantcharov et al (2007) The impact of objective assessment and constructive feedback on improvement of laparoscopic performance in the operating room. Surg Endosc, 21(12), 2240-3.						
To investigate the impact of assessment and constructive feedback on laparoscopic performance in the operating room	*Postgraduate* *Limited experience* 16 surgical trainees	*Psychomotor* *Operative* Laparoscopic cholecystectomy	*G1:* No feedback *G2:* 'Assessed' by experienced surgeon and detailed, constructive feedback provided	*Efficiency:* Time taken Economy of movement (via global rating scale) *Quality:* Number of errors (via global rating scale)	*Two procedure (two visit) RCT* Initial performance Intervention Repeat performance (within 2 weeks)	No inter-group baseline differences *Efficiency & quality:* G2 better in all three performance measures on retest
Porte et al (2007) Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. Am J Surg, 193(1), 105-10.						
To investigate the effect of different types of feedback on psychomotor skills	*Undergraduate* 45 medical students	*Psychomotor* *Bench model* Suturing and knot tying	*G1:* Computer-generated feedback re economy of movement (number of movements, time per movement, speed of movements, instrument total distance) *G2:* As above plus expert reference values *G3:* Verbal feedback from expert (summary feedback, after each trial)	*Efficiency:* Number of movements, time per movement, speed of movements, instrument total distance *Quality:* Performance on global rating scale	*Two visit RCT* Instructional video Pre-intervention performance 1-hour practice session (19 task trials) Immediate post-intervention performance. *Delayed retention performance 1-month post intervention (5 suture placement)*	*Efficiency:* All groups improved pre- to post-testing. *Quality:* Global rating scale (All groups improved pre- to post-testing) Only *G3* retained improvement at delayed testing
Rafiq et al (2008) Objective assessment of training surgical skills using simulated tissue interface with real-time feedback. J Surg Educ, 65(4), 270-4.						
To investigate the effect of real-time performance feedback on grip force whilst suturing	*Undergraduate* 12 medical students	*Psychomotor* *Bench model* Suturing (task 1: interrupted task 2: continuous, for 60 seconds)	*G1:* No feedback *G2:* Grip force data via graphic display – grip on forceps, suture holder and downward force on suture pad	*Quality:* Grip pressure (forceps, suture holder and suture pad)	*Single visit RCT* 15-minute training tutorial 15-minute practice Pre-intervention performance (task one and two performed in random order; 1 then 2, 2 then 1) Post-intervention performance: Tasks repeated with intervention	*Quality:* G2 better able to regulate pressure

Aim	Participants	Task	Intervention	Task measures	Design	Results
O'Connor et al (2008) How much feedback is necessary for learning to suture? Surg Endosc, 22(7), 1614-9.						
To investigate the effect of knowledge of results (KR) and knowledge of performance (KP) on laparoscopic skill acquisition and perception of work-load	*Undergraduate* *Novice* 9 medical students	*Psychomotor* *Virtual reality simulated laparoscopic task* Suturing and knot tying	*G1:* No feedback *G2:* Measures of performance on simulator (KR) *G3:* As above and instruction; detailed explanation, use of performance goals via instructor (KR & KP)	*Efficiency:* Task time, instrument path length *Quality:* Smoothness of instrument movement via expert global rating score	*Multiple visit RCT* 2-hour initial training session on hand instrument suturing and knot tying. Laparoscopic training session and 30 min practice. 1-hour training session, 6 days a week, for 4 consecutive weeks. Last trial in each session recorded as performance for analysis.	*Efficiency & quality:* Variation greater in *G1* than *G2&3*. *Efficiency: G2 & 3* better than *G1* in throughout study. *Quality: G3* made fewer knot-tying errors than *G2* or *G1*. All participants reached a plateau in performance by day 8. No need for such a long study.
Kruglikova et al (2010) The impact of constructive feedback on training in gastrointestinal endoscopy using high-fidelity Virtual-Reality simulation: a randomised controlled trial. Gut, 59(2), 181-5.						
To assess the impact of external feedback on colonoscopy simulator performance	*Postgraduate* *Novice* 22 clinical trainees	*Psychomotor* *Simulated colonoscopy trainer* Task 1: Simulated colonoscopy Task 2: Reduction of bowel loop	*G1:* No feedback *G2:* Structured feedback from experienced supervisor (concurrent and summary)	*Efficiency:* Procedure time *Quality:* Frequency of bowel perforation, % mucosa seen, volume of insufflated air	*Multiple visit RCT* Pre-trial video instruction Expert demonstration 15 repetitions of task 1 (over 4 sessions within 4 weeks, max 2 sessions/week, 3-5 repetitions/session) *Delayed testing:* Repeat performance of task 1 and 2; 4-6 weeks post last repetition	*Efficiency:* Both groups improved time efficiency. *Quality: G2* saw more of mucosa. In delayed testing, *G2* performed better in relation to time efficiency and % mucosa seen.

Aim	Participants	Task	Intervention	Task measures	Design	Results
El Saadawi et al (2010) Factors affecting feeling-of-knowing in a medical intelligent tutoring system: the role of immediate feedback as a metacognitive scaffold. Adv Health Sci Educ Theory Pract, 15(1), 9-30.						
To investigate if immediate feedback as a metacognitive scaffold and whether other forms of metacognitive scaffolding sustain performance during fading of immediate feedback?	*Postgraduate* 23 pathology residents	*Academic* Use of a pathology Intelligent Tutoring System (ITS) to work through virtual cases	*G1:* Immediate summary feedback post-clinical case *G2:* Immediate feedback plus 3 supplementary metacognitive scaffolds (which check confidence in correctness at stages during task)	*Quality:* Correlation between feeling of knowing and correctness in 3 tests (each with 4 cases)	*Two visit repeated measures RCT* Participants randomised between two groups 20-minute demo video of ITS followed by supervised practice in using the system. Pre-intervention performance Visit one: Both groups complete cases with immediate feedback. Intra-intervention performance Visit two: Both groups complete cases with intervention conditions Post-intervention performance	*Quality:* Intra-intervention performance: significant improvement in both group's correlation between FOK and correctness Post-intervention performance Correlation in both groups decreased. This was not protected by scaffolding in *G2*.
Boyle et al (2011a) Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure. Postgrad Med J, 87(1030), 524-8.						
To investigate the effect of feedback on laparoscopic task performance	*Postgraduate* 28 surgical trainees	*Psychomotor* *Virtual reality simulated laparoscopic task* 5 simulated laparoscopic colectomy procedures	*G1:* Able to ask questions to facilitators re instruments or procedure but no performance feedback *G2*: Self-assesses after each procedure. 'Standardised' facilitator feedback (no explanation of this) and given computer metrics.	*Efficiency:* Simulator generated instrument path length *Quality:* Simulator generated instrument path smoothness Expert calculated frequency of intra-procedure errors	*Single visit RCT* Standardised teaching: demo video by expert, explanation of performance metrics and errors, instrument instructions, practice task. 3 procedures performed with intra-procedure simulator prompts 2 procedures performed without intra-procedure simulator prompts	*Efficiency: G2* worse in respect to instrument path length *Quality: G2* worse in respect to instrument path smoothness. *G2* had fewer procedure errors than *G1* *(Results may be confounded as those participants who completed the task better moved their instruments further to do so)*

Aim	Participants	Task	Intervention	Task measures	Design	Results
Boyle et al (2011b) The importance of expert feedback during endovascular simulator training. J Vasc Surg, 54(1), 240-248 e1.						
To assess the importance of expert feedback during virtual reality simulator training	*Postgraduate Novice* 18 post-BST surgical trainees	*Psychomotor Simulated virtual reality task* Virtual reality simulated renal artery stenting	*G1:* No feedback; aware of duration of task *G2:* Non-expert (facilitator feedback), discussion of errors, VR-generated measures of performance *G3:* Expert feedback (summary)	*Efficiency:* Procedure time Volume of contrast used *Quality:* Accuracy of balloon placement, handling errors, procedural errors (expert scored)	*Single visit RCT* Didactic teaching of procedure in standard fashion (PowerPoint, demo, questionnaire) 6 task performances	*Efficiency:* no significant differences in most metrics although trend of *G3* better than *G2, G2* better than *G1*. *Quality: G3* made fewer handling errors; *G2* & *G3* made fewer procedural errors
Li et al (2011) Pre-training evaluation and feedback improve medical students' skills in basic life support. Med Teach, 33(10), e549-55.						
To investigate the impact of pre-training feedback on medical student performance in BLS	*Undergraduate* 40 medical students	*Academic* Written examination score *Skills task* Single BLS skills station	*All interventions pre-BLS course and testing* *G1*: Control; 45-minute BLS lecture *G2:* Individual mock arrest moulage, group feedback and 30-minute BLS lecture	*Quality:* Score in written exam Skills station score via examiner-scored structured checklist	*Single visit RCT* *Sample size calculation performed* Groups randomised Pre-course intervention BLS training Exam and task performance.	*Quality:* No difference in exam score. *G2* better in skills station.
Kannappan et al (2012) The effect of positive and negative verbal feedback on surgical skills performance and motivation. J Surg Educ, 69(6), 798-801.						
The effect of positive and negative verbal feedback on skill performance	*Undergraduate* 25 medical students	*Psychomotor Simulated laparoscopic task* Peg transfer task on laparoscopic box trainer	Individual participant feedback session from expert who did **not** watch the performance but were thought to have watched *G1:* General compliments *G2:* General critical comments	*Efficiency:* Time to complete the task *Quality:* Number of errors (sum of pegs dropped and failure to transfer between graspers)	*Single visit RCT* Instructional video Task practice 1 x pre-intervention performance Intervention 1 x post-intervention performance	*Efficiency:* Both groups improved significantly; *G2* tended towards greater improvement but not significant *Quality:* Both groups improved

Aim	Participants	Task	Intervention	Task measures	Design	Results
Aronson et al (2012) A comparison of two methods of teaching reflective ability in Year 3 medical students. Med Educ, 46(8), 807-14.						
To assess the impact of critical reflection guidelines, feedback re reflection, and interaction between the two on student's reflective ability in written reflections	*Undergraduate* 149 medical students	*Academic* 3 x pieces of coursework, including a reflective element	*1st intervention* *G1:* Given reflection guidelines inc. study exercise *G2:* No reflection guidelines *2nd intervention* *G1.1* and *G2.1:* Reflective content and process feedback *G1.2* and *G2.2:* Reflective content feedback only	*Quality:* Score of reflective ability in each submitted piece of work, graded by blinded faculty member	*Repeated measures RCT* All participants given definitions for reflection/critical reflection First intervention Submission of 3 pieces of work including reflective element with subsequent second intervention	*Quality:* No inter-group difference. Both reflective guidelines and process FB improved reflective performance
Wojcikowski & Kirk (2013) Immediate detailed feedback to test-enhanced learning: an effective online educational tool. Med Teach, 35(11), 915-9.						
To determine whether detailed post-question feedback enhanced learning	*Undergraduate* 103 clinical science students	*Academic* 20 multiple choice questions as part of end of year exam	*G1:* Answer only *G2:* Answer and detailed explanation re correct and incorrect answers	*Quality:* Score in end of year exam	*Two visit RCT over 2 years* Year one: *G1* Year two: *G2* Clinical case scenarios with 5 questions and 'feedback' as per intervention group. No mention of how many case scenarios done. End of year exam	*Quality: G2* performed better in end of year exam

Aim	Participants	Task	Intervention	Task measures	Design	Results
Strandbygaard et al (2013) Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial. Ann Surg, 257(5), 839-44.						
To investigate the impact of instructor feedback on performance of a complex simulated task	*Undergraduate* *Novice* 98 senior medical students	*Psychomotor* *Virtual reality simulated laparoscopic task* Laparoscopic salpingectomy (LapSim model)	*G1:* No feedback *G2:* One obligatory 'feedback' session with expert, option for two further sessions. Sessions were not individualised or based on observation or performance data but included instructions on completion of task	*Efficiency:* Total time and number of task repetitions *Quality:* Simulator-generated performance score	*Multiple visits over 2 months* *Sample size calculation* Induction meeting with explanation of task, instructions of use of model and of the automated performance data produced by LapSim model after task performance. Participant-controlled repeated visits and task performances until pre-defined level of task competence achieved, twice, within 5 repetitions (completion of task), with instructor 'feedback' sessions for participants in *G2*	Small number dropped out of control group due to frustration *Efficiency: G2* needed fewer task performances and less intra-task time to achieve completion *Quality:* Simulator-generated performance scores were higher in *G2*
Farjad et al (2013) Effect of feedback content on novices' learning ultrasound guided interventional procedures. Minerva Anestesiol, 79(11), 1269-80.						
To compare the effect of two forms of feedback on acquisition and retention of practical skill	*Undergraduate* 30 medical students	*Psychomotor* *Bench model* USS-guided aspiration with 5-step guidance given to standardise technique	*G1:* No feedback *G2:* Knowledge of results (KR); given measures of time taken and number of needle passes required at end of each task *G3:* Knowledge of performance (KP); *standardised* teaching on common errors and correction of technique via pre-recorded video at end of each task	*Efficiency:* Imaging time, needling time and total performance time *Quality:* Investigator calculated procedural error count via checklist	*Two visit RCT* Visit one: Standard training video (60 mins access) 5 x performance of task with intervention (participants asked to minimize imaging time, needling time and total performance time) Visit two: Task repeated once	*Efficiency:* Rapid improvement in all 3 groups. No inter-group differences *Quality: G3* better than *G2*, better than *G1*

Aim	Participants	Task	Intervention	Task measures	Design	Results
Farquharson et al (2013) Randomized trial of the effect of video feedback on the acquisition of surgical skills. Br J Surg, 100(11), 1448-53.						
To compare the effects of verbal vs verbal plus video feedback on psychomotor tasks	*Undergraduate* 48 medical students	*Psychomotor* *Bench model* Wound suturing	*G1*: Individualised summary feedback but no video footage to review *G2:* Individualised summary feedback plus video footage available to view at home afterwards	*Quality:* Investigator score via structured checklist and global rating score	*Two visit blinded RCT* Video demonstrating task Pre-intervention performance and feedback intervention Post-intervention performance (repetition of task at 24 hours)	*Quality: G2* were significantly better than *G1*, both in overall score and individual skill components
Paschold et al (2014) Tailored instructor feedback leads to more effective virtual-reality laparoscopic training. Surg Endosc, 28(3), 967-73.						
Do laparoscopic novices with lower initial performance scores benefit from intensive instructor feedback?	*Undergraduate* 20 medical students	*Psychomotor* *Virtual reality simulated laparoscopic task* Laparoscopic clip applying task (LapSim model)	*G1:* High performers, no feedback *G2:* Low performers, simulator-generated measures of performance (time taken, instrument path length, number of errors made) and 1:1 summary instructor feedback	*Efficiency:* Time taken to complete task, instrument path length *Quality:* Number of errors	*Single visit RCT* Task intro and tutor demo Pre-intervention performance: x 2; Low performers identified via lower than median score on initial task performance Training & intervention phase: completion of 6 training LapSim tasks Post-intervention performance: 3 x task performance	*Efficiency & quality:* No inter-group difference in post-intervention performance scores (significant difference present prior to intervention)

2.3.1 Study aims and concepts

The aims of this group of studies were to investigate, explore and evaluate the effects of feedback – in varying types and from various sources – on the performance of practical and academic tasks. Of these 22 interventional studies, eight studies employed a design which compared performance in one group, who received feedback, with the performance of a control group, who did not receive feedback. In nine studies, the effect of different types of feedback was compared; an additional control group featured in five of these nine studies (Table 2).

Table 2: Breakdown of study design by feedback intervention

Feedback vs Control	Feedback vs Feedback	Feedback vs Feedback vs Control
Rogers (2000)	Porte (2007)	Backstein (2004)
Boehler (2006)	El Saadawi (2010)	Judkins (2006)
Grantcharov (2007)	Aronson (2012)	Xeroulis (2007)
Kruglikova (2010)	Farquharson (2013)	O'Connor (2008)
Boyle (2011)		Farjad (2013)
Boyle (2011)		
Li (2011)		
Paschold (2014)		

In relation to the remaining five studies, there are significant issues with the 'feedback interventions' employed, limiting the validity of the conclusions and their relevance to the feedback literature. Kannappan et al (2012) purported to investigate the effect of positive versus negative expert feedback on learner psychomotor task performance. However, participants in both arms of the study received only generalised and generic *comments* from the expert instructors following task performance. Whilst one group received generalised compliments and the other negative remarks, these comments were unrelated in any way to the performance of the task. Therefore, whilst the results may be a basis for comment on the effect of encouragement and discouragement of learners, they cannot be used to comment on the effect of feedback. The 'feedback' interventions used by Wojcikowski and Kirk (2013) and Strandbygaard et al (2013) are representative of teaching and instruction rather than feedback. In Wojcikowski and Kirk's study, the group forming the experimental arm of the study received greater standardised explanation of written test answers during the learning phase. In Strandbygaard et al's study the experimental arm had greater access to non-individualised tutor task instructions. Therefore, these studies measure the effect of greater participant instruction on task performance, but not feedback. Finally, the ‘feedback’ interventions employed by van Sickle et al (2007) and Rafiq et al (2008) are more accurately described as real-time performance monitoring, than educational feedback. Van Sickle’s study compared the number of errors made during a psychomotor task when performed by a control group – who received no intra-performance notification of error – with three experimental groups, all of whom received a real-time audio notification when an error was made. Similarly, Rafiq’s study reported on the pressure exerted on a bench model during task performance by a control group – who had no ability to objectively measure pressure – with that of a group who were informed in real-time, via a graphic display, of the pressures they were exerting during task performance. Therefore, these studies may comment on the effect of real-time performance monitoring, but not the effect of educational feedback.

Two studies went beyond testing the effect of the presence versus absence of feedback, or the relative effect of different types of feedback. El Saadawi et al (2010) used a computer-based educational resource to investigate whether participants provided with greater access to teaching techniques that encouraged the development of metacognitive scaffolds during learning were able to retain information and apply knowledge to future performance. Aronson et al (2012) investigated the effect of explicit and implicit encouragement of self-reflection skills during feedback, on the self-reflection abilities of undergraduate students.

2.3.2 Study design

All 22 studies employed a quantitative design but only two (Li et al, 2011; Strandbygaard et al, 2013) included a sample size calculation in their design. This limits the collective quality of this pool of research. Twenty studies confined participants to two or more study groups via randomisation at the start of the study. Two studies featured a cross-over design: in Backstein et al’s experimental study (2004) participants rotated between study arms, completing one of the four psychomotor tasks under each feedback condition: Aronson et al (2012) introduced two different interventions at two different points in the study to produce four different study conditions. Interestingly, the two studies featuring a cross-over design, failed to demonstrate a significant inter-group quantitative difference in participant performance.

Eleven of the 22 studies were single-visit studies, with the effect of feedback measured during, and limited to, only one episode of contact with the participant. Seven studies employed a two-visit design, in which two separate visits were required in order to capture the post-intervention data. The duration of these studies were spread over a wide spectrum: from only 24 hours between study visits (Farquharson et al, 2013) and a full academic year (Wojcikowski & Kirk, 2013). An additional four studies employed a multiple-visits design, in which skills or knowledge was accumulated over greater study periods and multiple separate study visits. Again, the duration of these studies was spread over a wide spectrum: from four weeks (O'Connor et al, 2008; Kruglikova et al, 2010) to a full academic year (Aronson et al, 2012).

Three studies included an additional study visit to assess delayed performance and skill or knowledge retention. Xeroulis et al (2007) and Porte et al (2007) assessed the efficiency and quality of participants' suturing and knot tying four weeks after initial post-intervention measurement. Kruglikova et al (2010) re-assessed participants' abilities 4-6 weeks post initial post-intervention measurement in relation to the primary task, of colonoscopy on a virtual-reality simulator model, and also their skill in relation to a second task on the simulator (reduction of a bowel loop).

2.3.3 Participants

The majority of these studies (n=16) recruited undergraduate medical students as participants. One recruited a combination of undergraduate medical and science students. In four of these studies, control of experience was explicitly mentioned, with 'novice' learners (in relation to the task involved) specifically recruited. In the remaining six studies, post-graduate training-grade doctors from various clinical specialties were recruited, with participants recruited from only one specific specialty in each study. In three of these studies, control of experience was explicitly mentioned, with 'novice' learners (in relation to the task involved) specifically recruited. None of the studies recruited non-training grade clinicians.

2.3.4 Task selection

Only five of the studies took advantage of learning situations that naturally occurred as part of undergraduate or postgraduate training. Three of these studies (El Saadawi et al, 2010; Aronson et al, 2012; Wojcikowski et al, 2013) focused their investigation of feedback in relation to an academic task (postgraduate pathology case interpretation, undergraduate essay submission and an undergraduate summative exam result respectively). Li et al (2011) utilised a Basic Life Support (BLS) course, and alterations to the teaching and feedback received by a subgroup of candidates and measured the effect of feedback via an examiner-marked skills station and post-course written test score. Grantcharov et al (2007) based their study in clinical practice and measured the effect of feedback given by experienced surgeons on the efficiency and quality of the participant performance of a psychomotor task. They recruited 16 surgical trainees of ‘limited experience’ and assessed their performance during two laparoscopic cholecystectomies, having provided one group with ‘detailed, constructive feedback’ and the other with no feedback.

The majority of the studies (n=17) conducted their investigations in a simulated environment, measuring the effect of feedback via the performance of psychomotor tasks. Table 3 illustrates the breakdown of these psychomotor tasks by model used and task performed. A ‘bench model’ describes a physical model that creates a simulated opportunity for a medical or surgical procedure, typically using hand instruments in combination with synthetic or animal models. Laparoscopic simulators can be divided into two main types: a ‘box training model’, in which hand instruments are inserted into a physical box, viewed via a screen connected to a portable camera placed within the box, in which a physical task is carried out; or a ‘virtual reality laparoscopic simulator’, in which specially designed hand pieces are connected to a PC and simulated tasks are completed via a software programme. ‘Task-specific model simulators’ also utilise virtual reality and simulation software. These complex models are designed in relation to one specific task, such as the colonoscopy trainer utilised by Kruglikova et al (2010).

Table 3: Breakdown of psychomotor tasks utilised, by model and task performed

Type of model used	Author	Task(s) performed
Bench model	Rogers (2000)	*Surgical knot tying*
	Backstein (2004)	*Sawbone plating, fixation of olecranon, Z-plasty (porcine skin model)*
	Boehler (2006)	*Surgical knot tying*
	Xeroulis (2007)	*Surgical knot tying (hand and instrument tying)*
	Porte (2007)	*Surgical knot tying*
	Rafiq (2008)	*Surgical knot tying*
	Farjad (2013)	*USS-guided aspiration and injection*
	Farquharson (2013)	*Wound suturing*
Virtual reality laparoscopic simulator	Judkins (2006)	*Bimanual object carry, needle passing and suture tying*
	O'Connor (2008)	*Laparoscopic suturing and knot tying*
	Boyle (2011)	*Laparoscopic cholecystectomy*
	Boyle (2011)	*Renal artery stenting*
	Strandbygaard (2013)	*Laparoscopic salpingectomy*
	Paschold (2014)	*Laparoscopic clip applying task*
Laparoscopic box training model	Van Sickle (2007)	*Maze tracking (practice), cutting task (incisions in piece of paper; testing)*
	Kannappan (2012)	*Peg transfer task*
Task-specific model simulators	Kruglikova (2010)	*Colonoscopy and bowel loop reduction*

2.3.5 Measuring the effect of feedback

Analysis of the methodologies employed in this subsection of the medical and surgical education literature illustrates that all studies focused on the measured quantified effect of feedback. These studies aimed to assess how feedback might affect task performance but do not explore why or how this might be the case. Performance data was obtained both objectively, via computer measurements (time taken, instrument path length) or binary count (the number of errors made), or subjectively, such as via expert observer-generated global rating scale scores. The performance data collected can be categorised as measures of efficiency or of quality.

The measurement of performance with academic tasks was obtained via scores attained on written papers and assignments and, therefore, these studies focused only on the quality of performance. They were unable to measure efficacy of participant performance. Li et al (2011) objectively measured the effect of feedback on BLS course participant performance via their post-course test scores and subjectively via the examiner-generated post-course skills station score. El Saadawi et al (2010) and Wojcikowski et al (2013) both measured quality of performance objectively, via statistical correlation between perception and actual correctness and written test scores respectively. Aronson et al (2012) obtained subjective quality of performance data via faculty grading of participant reflective ability.

When examining the studies featuring performance of a psychomotor task, there were three main groups in relation to measures of performance: those that employed both objective and subjective measures to assess both efficiency and quality of performance; those that employed only objective measures of both efficiency and quality of performance; and those that employed subjective measures only to assess quality of performance only.

2.3.5.1 Studies utilising subjective and objective measures of efficiency and quality of performance

Six studies (Rogers et al 2000; Xeroulis et al 2007; Grantcharov et al, 2007; Porte et al, 2007; O'Connor et al, 2008; Boyle et al, 2011) fall into this category of measures of task analysis. These studies tended to use objective measures of performance, such as time taken to complete the task and path length of instruments, to measure the efficiency of performance (likely due to ease of collection of this data) but employed subjective measures in relation to quality of performance. Quality of performance was measured by expert-generated global scores, evaluating domains such as smoothness of movement, accuracy of instrument placement or overall competence. In these studies, owing to the study design and model used and task completed, these measures were more difficult to objectively quantify.

2.3.5.2 Studies utilising only objective measures of efficiency and quality of performance

Eight studies (Judkins et al, 2006; van Sickle et al, 2007; Kruglikova et al, 2010; Boyle et al, 2011; Kannappan et al, 2012; Strandbygaard et al, 2013; Farjad et al, 2013; Paschold et al, 2014) used only objective measures of efficiency and quality of performance. With the exception of Farjad et al (2013), all of these studies employed a design that included performance of a task on a virtual reality laparoscopic simulator. These machines automatically collect objective data relating to efficiency of performance - such as time taken to complete the task and path length of instruments - but also quality of performance via error count, such as number of pegs dropped or percentage of mucosa visualised.

2.3.5.3 Studies utilising only subjective measures of quality of performance

Three studies (Backstein et al, 2004; Boehler et al, 2006; Farquharson et al, 2013) employed expert observer-generated global rating scale scores of participant performance in their studies, with no objective quantified data collected. All three of these studies employed a bench model task, with resulting difficulty in collection of objective data. Whilst it might be argued that this type of assessment of performance most closely mimics the type of assessment encountered in clinical training, it lacks specificity and may be subject to observer bias.

2.3.6 The feedback intervention

The types of feedback intervention featured in this section of the literature can be considered by source, style and content. There are broadly three sources of feedback featured in the 22 studies: an 'expert' instructor, quantitative data from the simulator models used for task completion and participant self-reflection.

Feedback from an 'expert' instructor is a common feature of these studies (Table 1). This summary feedback consists of an individualised performance review based upon observed behaviour, with critique, praise, suggestions and explanation transmitted from the expert (tutor) to the participant (learner). The feedback described is expert driven and expert dependent and of the information transfer model. However, whilst some of the studies purport to employ expert-derived feedback as the feedback intervention, there are multiple studies identified in this literature review that involve teaching, with instruction given regarding a task rather than feedback based on observed behaviours. In addition to the work of Strandbygaard et al (2013) and Wojcikowski et al (2013), Boyle et al (2011) described that participants in one feedback intervention group could ask questions in relation to the task but did not receive performance feedback. Li et al (2011) used a mock-arrest scenario rather than a lecture in the intervention arm before testing performance in a Basic Life Support skills station. Rogers et al (2000) and Xeroulis et al (2007) utilised trainer instruction and discussion during the learning phase of their studies (concurrent feedback or coaching), rather than summative feedback based upon previous performance.

The feedback intervention in several studies involved the communication of quantitative data collected during task performance to participants. The most common form of this intervention was participants being informed of measures of performance such as time taken to complete a task, instrument path length or number of errors made, after task completion. However, others employed the real-time communication of quantitative data, such as grip strength (Rafiq et al, 2008) or notification of error (van Sickle et al, 2007) as feedback. In these studies, the feedback model is more akin to the biological than educational model, with performance measures regulating concurrent performance.

Backstein et al (2004), Aronson et al (2012) and Farquharson et al (2013) incorporated elements of participant self-reflection into their feedback interventions. In the studies measuring the performance of a psychomotor task, this self-reflection was invited via participant review of their recorded task performance but no instruction was given in relation to how this review should be conducted.

2.3.7 Study results

In relation to the effect of feedback on task performance, these studies show varied results (Table 4). Eight of the 17 studies analysed showed a positive effect associated with feedback; five studies illustrated mixed results (with positive effects associated with some measures of performance but no effect with others), and the remaining four studies failed to show any difference in task performance in relation to provision or absence of feedback.

Table 4: Study results; the effect of feedback in relation to intervention design and measure of performance

	Feedback vs Control studies	**Feedback vs Feedback studies**
Feedback associated with improved task performance	* Roger *(Quality, subjective)* * Li *(Quality, subjective)* * Kruglikova *(Quality, objective)* * Boyle II *(Quality, subjective)* Boehler *(Quality, subjective)* Grantcharov *(Efficiency, objective; (Quality, subjective)* Paschold *(Efficiency & Quality, objective)*	* Farjad *(Quality, subjective)* Judkins *(Efficiency & Quality, objective)* Xeroulis *(Efficiency, objective & Quality, subjective)* Porte *(Efficiency, objective & Quality, subjective)* O'Connor *(Efficiency, objective & Quality, subjective)* Farquharson *(Quality, subjective)*
Feedback associated with no difference in task performance	* Rogers *(Efficiency & Quality, objective)* * Li *(Quality, objective)* * Kruglikova *(Efficiency, objective)* * Boyle II *(Efficiency, objective)* El Saadawi *(Quality, objective)* Boyle I *(Efficiency, objective & Quality objective & subjective)*	* Farjad (*(Efficiency, objective)* Aronson *(Quality, objective)* Backstein *(Quality, subjective)*
* Results relating to efficiency and quality measures show different effects within the same study		

Within the five studies that showed a mixed effect of feedback, an interesting trend is evident: feedback tended to be associated with positive effects in relation to the quality of task performance when measured subjectively, but not with the efficiency of task performance when measured using objective measures (Table 4). Four of these five studies measured efficiency via time to complete the task. However, this pattern was not seen in the studies that found that feedback was associated with improved performance in relation to all measures of performance (including efficiency). Six out of these eight studies included objective measures of efficiency (rather than only subjective measures of quality (Boehler et al, 2006; Farquharson et al, 2013)); and five of these six included more than one objective measure of efficiency (instrument path length, speed of movement, number of hand movements made) rather than considering only time taken to complete the task. Therefore, whilst there is evidence within the literature to suggest that feedback is associated with improved performance, it also suggests that it is more difficult to prove this association via objective measures of efficiency than it is via subjective measures of quality.

When analysing the feedback interventions utilised in the eight studies showing only positive effects associated with feedback, six of the eight (Boehler et al, 2006; Grantcharov et al, 2007; Xeroulis et al, 2007; Porte et al, 2007; Farquharson et al, 2013; Paschold et al, 2014) employed 'expert' feedback. Xeroulis et al (2007) and Porte et al (2007) identified that summary expert feedback (provided *after* task performance) was associated with better performance at immediate post-intervention and delayed retention testing. The former illustrated that, unlike summary feedback, concurrent expert feedback (provided during each task performance) was not associated with improved delayed task performance, which may reflect cognitive over-load of the learner. All eight studies employed a design featuring completion of a psychomotor task. These results support the assertion that expert (information transfer) feedback can be a powerful tool for improving psychomotor task performance.

The only study to investigate the effect of participant self-reflection (via task performance video review) was Farquharson et al (2013). In this study, both arms of the study received individual summary feedback after task performance, but one group also received a copy of a video of their task performance for review out with the study session. The participants in this group were asked to watch the video and use a standardised proforma provided by the investigator to identify areas for improvement. No further details were given about the proforma or depth of participant analysis, although checks were put in place to ensure compliance with these instructions. Nor was it contemplated that consideration of the proforma itself might have been a cause for improvement in participant performance. However, this study found that the group that engaged in self-reflection and analysis out-performed those receiving only summary feedback on the post-intervention testing, as measured by a subjective measure of quality (investigator-generated global rating scale). The results of this study would support further investigation of learner self-reflection and its potential association with improved psychomotor task performance.

The four studies that failed to show any difference in task performance with feedback varied from one another in design. Two used a psychomotor task to measure the potential effect of feedback (Backstein et al, 2004; Boyle et al, 2011) and two featured an academic task (El Saadawi et al, 2010; Aronson et al, 2012). A mix of objective and subjective measures of performance were examined (Table 4). El Saadawi et al (2010) failed to show any difference in performance from participants who received greater metacognitive scaffolding resources during learning than participants that did not. Aronson et al (20102) failed to show any significant quantitative difference in the reflective abilities of those provided with self-reflection guidelines and feedback compared to those receiving only guidelines or feedback. Although Backstein et al s (2004) found no statistically significant difference in performance seen between participants asked to engage in self-analysis of previous performance (via video review) and those receiving expert summary feedback in addition to self-review, both of these groups tended to perform better than the control group. In fact, the group that engaged in self-analysis without expert feedback improved the most between pre- and post-intervention testing. Therefore, although the results were not statistically significant, this study echoes the later findings of Farquharson et al (2013) in relation to the positive effect of self-reflection on future performance.

2.3.8 Summary

This structured literature search identified 22 papers which investigate the effect of feedback via an experimental approach. The majority of the first named authors are clinicians and the work was commonly published in clinical journals. Most of the studies recruited undergraduate medical students as participants and, therefore, their findings relate to mostly novice learners. These studies were largely conducted in a simulated environment, with the advantages of increased control over study conditions and ease of quantitative data collection.

The feedback interventions employed can be classified into three groups: Expert-generated feedback (instruction, explanation, generalised encouragement or discouragement, or suggestions for change); the communication of quantitative performance data (akin to a biological model of feedback); or those incorporating elements of unstructured participant self-reflection (via review of performance videos). Despite the existence of this last, smaller subgroup of feedback intervention, the majority of studies adhered to an information-transfer model of feedback.

In general, the quality of quantitative design was poor, with only two studies featuring a sample size calculation in order to adequately power analysis. Half of the studies involved only one study visit, during which the feedback intervention was carried out and immediate effect measured. None of the studies featured a cross-over task. Only three studies incorporated a delayed measure of performance more than one week post intervention.

The effect of feedback was measured by both objective and subjective measures of efficiency and quality. These studies showed variation in the effect of feedback on task performance. Not all showed improved participant performance post feedback, with a positive effect more likely to be demonstrated when measuring quality of performance subjectively than quantitative performance objectively. A single study suggested that self-reflection without external feedback may be more effective.

2.4 DISCUSSION OF THE THEORY OF FEEDBACK IN THE MEDICAL AND SURGICAL EDUCATION LITERATURE

Via this systematic literature review, 24 papers were identified within medical and surgical education literature, in which a major focus was the discussion of the theory of feedback (Figure 4). One finding of note is that only one paper with this focus was identified within the surgical education literature (Sadideen & Kneebone, 2012). This contrasts with the 15 experimental studies previously identified and discussed from the surgical education literature, highlighting a paucity of theoretical studies based in surgical education. Resultantly, this discussion and analysis of the body of work pertaining to the theory of feedback is almost exclusively a representation of work published within the medical education literature.

2.4.1 Source, author and referencing of published work

Of the 24 articles, the majority (n=17) were published in medical education journals; six were published in clinical medical or surgical journals and one in an academic pan-health professional periodical. The first of the articles included was published in 1983 (Ende, 1983) and the most recent in 2015 (Kraut et al, 2015). Authors include clinicians, medical education academics and non-medical education academics. Within this body of work, a wide range of supportive literature is referenced: education, behavioural science, psychology, clinical studies, business administration and management theory, human factors and non-medical sciences.

2.4.2 Starting with Ende

Before further discussion of this body of work, it is worthwhile to set the feedback scene. Ende's seminal paper in 1983 not only represents the earliest work in this systematic review but the most often cited by other work. The didactic, information-transfer model of feedback presented in Ende's paper remains the dominant model to this day (as can be seen in the types of feedback highlighted above). It is this linear model around which all future discussion is set. The 'provision' of quality feedback by tutors and the investigation of how educators can promote learner 'compliance' with this feedback remains of interest and importance both to medical educators and to many clinical educators today.

It is worth noting that the emphasis Ende placed on the importance of encouraging participant understanding and insight as part of feedback was possibly ahead of its time; some years passed before medical education returned to that idea. Furthermore, his 'Guidelines for giving feedback' (Box 2), although general, would likely be readily accepted as best practice by most clinical educators in modern practice.

Box 2: Guidelines for giving feedback (Reproduced from Ende, 1983)
Feedback should be undertaken with the teacher and trainee working as allies, with common goals Feedback should be well-timed and expected Feedback should be based on first-hand data Feedback should be regulated in quantity and limited to behaviours that are remediable Feedback should be phrased in descriptive non-evaluative language Feedback should deal with specific performances, not generalisations Feedback should offer subjective data, labelled as such Feedback should deal with decisions and actions, rather than assumed intentions or interpretations

2.4.3 Study design and thematic analysis of published work

The design of the papers presented can be broadly categorised as review articles (n=4), presentations of original work (n=13), or commentary pieces (n=7). The dominant feature of a review article is the summary of previously published work without evidence of critique of the theories presented nor presentation of new theories (Wulf et al, 2010; Schartel, 2012; Sadideen & Kneebone, 2012; Kaul et al, 2014). An article was defined as original work if it was a complete journal article (not commentary) with evidence of synthesis of new ideas relating to the theory of feedback, evidence of application of theory to practice with conclusions drawn, or discussion of theory with resulting recommendations for practice (Ende, 1983; Rolfe & McPherson, 1995; Branch & Paranjape, 2002; Milan et al, 2006; van de Ridder et al, 2008; Sargeant et al, 2008; Bing-You & Trowbridge, 2009; Archer, 2010; Kluger & van Dijk, 2010; Watling et al, 2012; Rudland et al, 2013; Telio et al, 2015; Sandars, 2015). Finally, articles forming a commentary piece, and specifically published as such, in response to the work of others, made up a significant and interesting proportion of work pertaining to the theory of feedback in the medical education literature (Norcini, 2010; Molloy, 2010; Overeem, 2010; Murdoch-Eaton, 2012; Ajjawi, 2012; Archer, 2013; Kraut et al, 2015).

2.4.3.1 Review articles

The main collective theme of these articles is, as previously discussed, the presence of a summary, which is present in all four articles in this category. Wulf (2010), a non-medical scientist, presented a targeted review of feedback as an influential factor in respect to motor skill acquisition. Schartel (2012), an anaesthetist and Training Programme Director, provided a detailed summary of the theoretical underpinning of feedback practices in clinical training. Sadideen (Sadideen & Kneebone, 2012), a plastic surgery trainee, focused on theory regarding the tutelage of practical skills, referencing Fitz and Posner’s three-phase theory of motor skill attainment. Kaul (Kaul et al, 2014), a paediatrician, presented a review of published feedback theory specifically relating to the setting of paediatric and adolescent gynaecology.

Also present in all four papers were specific recommendations regarding feedback practice. Schartel (2012), and Kaul et al (2014), made recommendations regarding the specifics of feedback ‘delivery’ (timing, location, language used during and the quantity of feedback 'given'), and the principles underpinning 'effective' feedback (knowing your learner, understanding what is to be learned, creating a partnership) respectively. Sadideen & Kneebone (2012), advocated learning practical skills by volume, exposure and practice. Recommendations contained within review articles are the reflected recommendations of presented and referenced work, the often-condensed recommendations of other authors, not the presentation of original recommendations of best practice.

These summaries and recommendations are context and intended reader-specific. Three of these papers were published in clinical journals (Schartel, 2012; Sadideen & Kneebone, 2012; Kaul et al, 2014), reflecting the desire for medical and surgical subspecialties to present and read work relating to the theory of feedback in their own familiar, clinical context rather than in medical education journals.

2.4.3.2 Original works

The papers comprising this group can be further categorised by their dominant theme and, in this respect, lend themselves to separation into two groups: those dominated by opinion and those dominated by the presence of a combination of more complex cognition; the presentation of new ideas, critique of the existing body of relevant work, and discussion of concepts relating to feedback.

The four articles dominated by author opinion are detailed in Table 5. Ende's work in 1983 published the authors' clinically-orientated definition of feedback (p.777):

'*Feedback refers to the information describing students' or house officers' performance in a given activity that is intended to guide their future performance in that same or related activity'.*

The paper lays out clearly Ende's 'Guidelines for giving Feedback'. Although these recommendations are framed within the context of previous work from other sources, the style of the paper is editorial, with the authors' opinion on the paucity of, challenges to and best-practice in relation to feedback in the clinical domain the dominant feature.

Table 5: Opinion-dominated original work

Author	***Year***	***Journal***	***Title***
Ende	*1983*	*JAMA*	Feedback in clinical medical education
Branch	*2002*	*Acad Med*	Feedback and reflection: teaching methods for clinical setting
Bing-You	*2009*	*JAMA*	Why educators may be failing at feedback
Sandars	*2015*	*Int J Med Educ*	The challenge of feedback insights from non-medical education research

Similarly, Sandars' editorial (2009) relates the authors' opinion to a specific issue: the under-utilisation of non-medical education work by those in the medical education field. In this article the author selects references from the larger body of feedback literature to provide evidence that three important principles of feedback – that it should be part of a wider assessment for learning, that appreciating the feedback preferences of students is important, and that effective feedback approaches should be used – could be taken from non-medical education research and applied with benefit to medical education.

Branch and Paranjape (2002) adopted a conversational style in presenting the authors' duration-dependent categorisation of feedback episodes into brief, formal and major, and the subsequent specific recommendations relating to the role of the teacher in each. Bing-You & Trowbridge (2009) targeted the difficulties posed in relation to good feedback. They offered three specific reasons that physicians may be 'failing' at feedback (poor ability of learners for self-assessment, overpowering influence of affective reaction to feedback, and lack of adequately developed metacognitive capacities [of the learner]).

The other eight pieces of original work captured in this literature review are detailed in Table 6. The themes existing in these articles are more complex than the giving of opinion - the synthesis of new ideas, critique of the existing body of relevant work, and conceptual thinking in relation to feedback - and these often exist in combination. Interestingly and in contrast to the opinion-dominated work, these articles were all published in academic journals.

Table 6: Complex original work

Author	*Year*	*Journal*	*Title*
Milan	*2006*	*Teach Learn Med*	A model for educational feedback based on clinical communication skills strategies: beyond the "feedback sandwich"
van de Ridder	*2008*	*Med Educ*	What is feedback in clinical education?
Sargeant	*2008*	*J Contin Educ Health Prof*	"Directed" self-assessment: practice and feedback within a social context
Archer	*2010*	*Med Educ*	State of the science in health professional education: effective feedback
Kluger	*2010*	*Med Educ*	Feedback, the various tasks of the doctor, and the feedforward alternative
Watling	*2012*	*Med Educ*	Understanding responses to feedback: the potential and limitations of regulatory focus theory
Rudland	*2013*	*Clin Teach*	A student-centred feedback model for educators
Telio	*2015*	*Acad Med*	The "educational alliance" as a framework for reconceptualising feedback in medical education

The combination of these three themes is illustrated in Figure 5. This diagram illustrates that the main theme represented in this section of the literature is the presentation of new ideas. All but one of these papers (Archer, 2010) contains the proposals of new, original ideas, specifically relating to the understanding and theory of feedback in medical education.

The objective of Van de Ridder et al (2008) was to achieve a unifying definition of feedback in medical education. The authors employed a structured literature search methodology, exploring texts in the scientific, social science and medical education literatures. Their assertion, therefore, that the pre-existing definitions of feedback in medical education are based upon the concepts of information and reaction, and the subsequent definition proposed, are well-supported by the literature but not new to the literature.

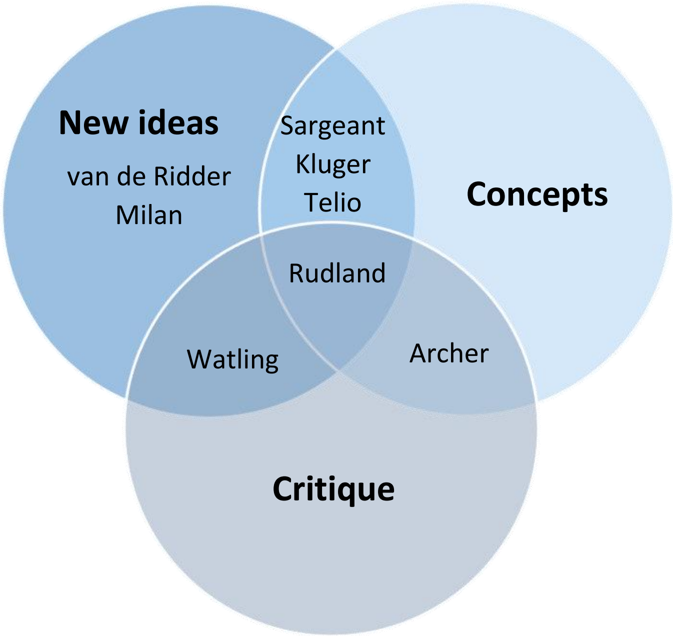
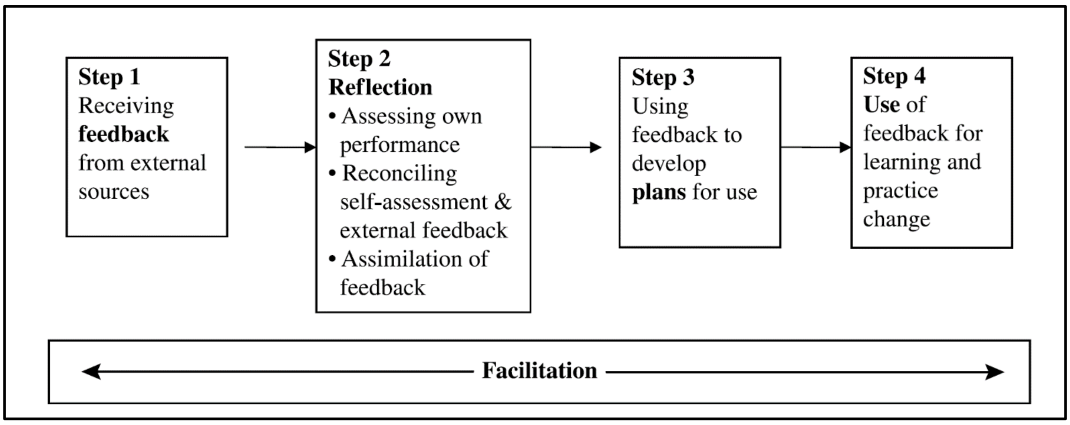


Figure 5:
Complex original work: the combination of themes by author

Sargeant et al (2008), Kluger and van Dijk (2010) and Telio et al (2015) engaged with analysis of concept in their respective publications and presented *new conceptual ideas*. Sargeant et al (2008) presented the concept of and model for 'directed' self -assessment. Drawing on an extensive body of personal research (largely based upon self-assessment rather than feedback), Sargeant proposed this model of interaction between feedback reception, self-reflection and assimilation, developing plans to use feedback and using feedback for practical change (Figure 6). Two important observations can be made: firstly, that aside from educator 'facilitation', there is no exploration of how this process works or how it can be promoted. Secondly, despite the model being termed ''directed' self -assessment'; the main subject of each stage of the model is feedback, hence its inclusion in this literature review.

Figure 6: Sargeants' model of 'directed' self-assessment (Reproduced from Sargeant et al, 2008)

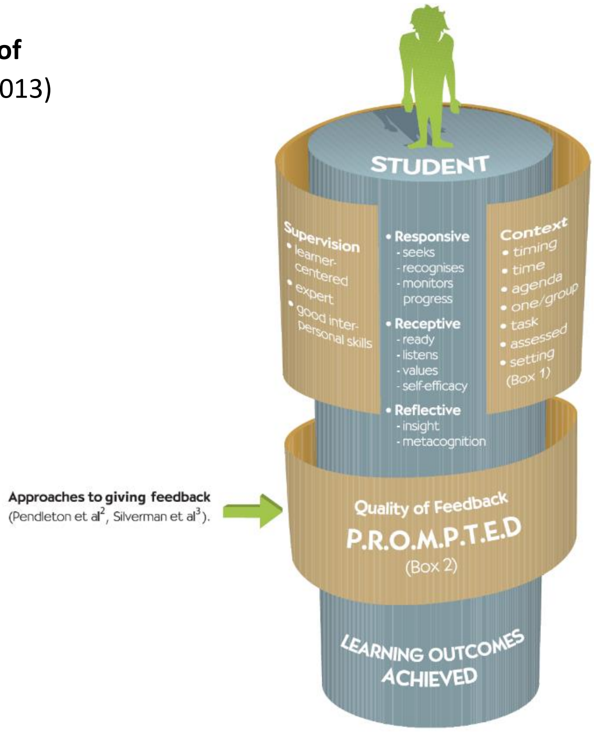


Kluger and van Dijk (2010) applied the social psychology principle of self-regulatory states (promotion and prevention) and asserted that positive or negative feedback would be received by the learner in different ways depending on the task and the subsequent regulatory state adopted. This work proposes that when adopting the prevention state (associated with mandatory tasks), negative feedback is more motivating and highly-regarded by the learner than positive feedback. Conversely, when adopting the promotion state (associated with non-mandatory tasks, which the learner undertakes in pursuit of some sort of gain), negative feedback is likely to demotivate and cause disengagement with the task but positive feedback is motivating. This work was further investigated by Watling et al (2012). This work supported the hypothesis proposed by Kluger and van Dijk (2010) but expanded upon some of the complexities of application of self-regulatory theory to feedback: that the regulatory focus adopted in relation to a task can be mixed (elements of both positive and negative) and that other factors - such as source credibility and content – can affect the acceptability of feedback.

Based on self-regulatory theory, Kluger and van Dijk (2010) proposed the *feedforward interview (FFI)*. The theory of the FFI is based upon modification of part of the appreciative enquiry theory used in the corporate business performance appraisals. Appreciative inquiry is a strengths-based approach that seeks to build on participants' experiences and aspirations. The theoretical advantage of the FFI is that by substituting the need for negative external feedback to be given, and replacing it with asking learners to focus on a behaviour that brought them reward, it removes the barrier between negative feedback and learner assimilation.

Telio et al (2015) contributed a piece of work which is both rich in its conceptual analysis of feedback in medical education and contributes new ideas to this topic. In this paper, the authors recognised the difficulties with feedback (differing perceptions, the emotive aspects, source credibility) but rather than support persistence with best practice recommendations designed to negate these issues, they call for a reconceptualisation of feedback itself. They explore and advocate the application of the psychotherapy phenomenon of the therapeutic alliance to the practice of medical education through the creation of the educational alliance.

Figure 7: Rudland's student-centred model of feedback (Reproduced from Rudland et al, 2013)

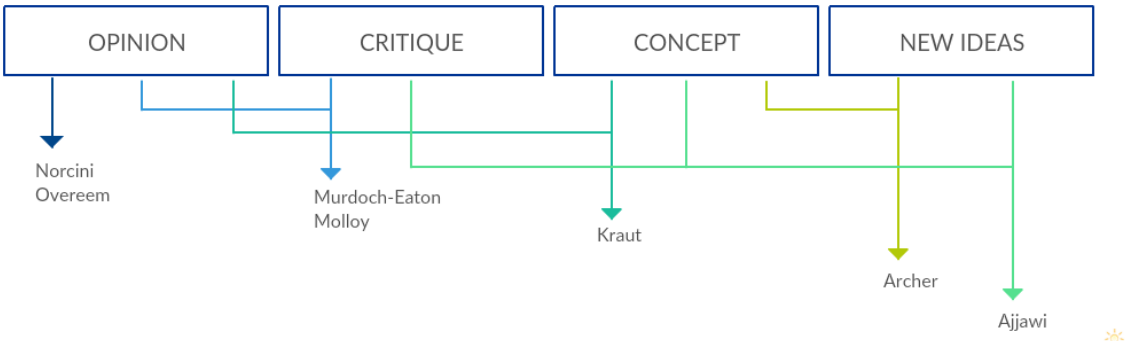


Similarly, Archer's paper 'State of the science in health professional education: effective feedback' (2010) goes beyond summary of previously presented evidence and utilises critique to challenge the reader's conceptualisation of feedback in medical education. Archer asserts that the currently employed models of feedback (the 'feedback sandwich' and Pendleton's model) are 'reductionist' in approach and the role of the learner is limited by an educator-driven process. This work supports a new 'model' for feedback: a change to the 'culture of feedback', in which learner self-monitoring is encouraged and informed by external feedback, and creation of a 'feedback continuum', in which 'feedback is reconceptualised as a supported sequential process rather than a series of unrelated events' (p.106). The limitation of this work is that, whilst conceptually powerful, the practicalities of implementing such changes are not addressed. Consequently, the 'model' is less of a fully-formed schema, and more of an outline of principle.

Finally, via targeted summary of existing evidence, concise and authoritative critique, and challenge to the concepts surrounding feedback, Rudland et al (2013) contributed a new 'student-centred' model of feedback (Figure 7). The authors assert that the tutor-centred approach featured in other existing models of feedback (Silverman, 1996; Pendleton, 2002) and the under-playing of the role of the learner 'undermines the objective' of feedback. In this feedback model, Rudland et al emphasise the 'main' and active role of the student, who must seek and respond to feedback. They assert that feedback must serve to improve student insight and that for successful feedback to occur, students must be motivated to learn, adopt better learning strategies and respond to situational demands. Their eight characteristics of quality feedback are detailed, and include it being precise, descriptive, and encouraging and constructive. However, they also suggest that feedback should be measurable and outcome based, which emphasises the observed effect of feedback on performance and is slightly at odds with the stated focus of learner growth.

2.4.3.3 Commentary articles

Figure 8: Commentary articles: the combination of themes by author



Commentray articles form a significant proportion of the literature reviewed (n = 7). The dominent themes represented in these papers are similar to those in 'original work' with the addition of opinion. A graphical illustration of the combined presence of these themes is given in Figure 8. These short pieces, published in the 'commentaries' or 'letters to the editor' sections of the educational journals, offer the opportunity for the authors to offer targetted comments and ideas, without the burden of the comprehensive feedback summary necessitated by a review article or presentation of original work. They offer the potential for prompt communication and exchange of ideas within medical academia. This type of work emerges towards the latter end of the literature review period (2010 – 2015). Summary details of this group of articles are given in Table 7.

Table 7: Commentary articles

Author	***Year***	***Journal***	***Title***
Norcini	*2010*	*Med Educ*	The power of feedback
Molloy	*2010*	*Med Educ*	The feedforward mechanism: a way forward in clinical learning?
Overeem	*2010*	*Med Educ*	'Paying it forward': performance improvement through feedforward interviews
Murdoch-Eaton	*2012*	*Med Educ*	Feedback: the complexity of self-perception and the transition from 'transmit' to 'received and understood'
Ajjawi	*2012*	*Med Educ*	Going beyond 'received and understood' as a way of conceptualising feedback
Archer	*2013*	*Med Educ*	Feedback: it's all in the CHAT
Kraut	*2015*	*J Grad Med Educ*	Feedback: Cultivating a positive culture

Murdoch-Eaton (2012) reflected on the limitation of retrospective feedback studies, and the complexities of the internal and external influences on the learner during the feedback process. She explored the contradictory nature of self-consistency theory (the learner's desire for predictable treatment) and self-enhancement theory (a learner's desire to increase self-worth) and the difficulty, therefore, with applying them to feedback in a practical manner. Murdoch-Eaton cited Watling et al's work in relation to regulatory theory (2010) as offering the tutor insight into how learner responses may be understood. Molloy (2010) goes one step further in this respect and suggests that when the tutor is able to recognise the learner's self-regulation focus, promotion or prevention, this can inform the tutor how to effectively tailor the type of feedback given to maximise effect.

However, Ajjawi (2012) challenges Murdoch-Eaton's assertion that a reconceptualisation of feedback should focus on a change from *'transmit'* to *'received and understood'*. Ajjawi asserts that this is not conceptually far enough. In this letter to the editor of the Medical Education, she focuses on the power of language; that *'received and understood'* perpetuates a *'narrow, transmissive view of learning'* (p.1018) that neither provides space for communication dialogue nor the requirement for the learner's role to develop and promote self-regulation.

Kraut et al (2015) further recognised the paradigm shift occurring in medical education, from an instructor-focused delivery of feedback to a learner-focused model that seeks to understand feedback and nurture 'reflection-in-action'. Unfortunately, whilst the authors urge the medical education community to re-examine current feedback practice, by their own admission, they could not understand nor articulate yet, how this new model might be facilitated.

2.4.4 Assessment or feedback?

Within this body of work there is discord in relation to the proposed relationship between feedback and assessment. Two papers advocate separation of feedback (information) and evaluation (judgement) (Ende, 1983), and information given regarding success or otherwise in relation to a goal (assessment) and feedback (Branch & Paranjape, 2002). Yet van de Ridder et al (2008), who sought to define feedback in medical education, was published in the 'assessment' section of the medical education journal in question.

Sandars (2015) defined formative assessment as 'assessment for learning'; a view supported by Schartel (2012), describing feedback, measuring performance compared to goals, as part of formative assessment. Sargeant et al's proposed model for 'directed' self-assessment (2008) relied upon the input of external feedback. Assessment forms part of the contextual issues that affect feedback in Rudland et al's (2015) new model of student-centred feedback (Figure 7). In this model, assessment and feedback are said to be 'intertwined'.

2.4.5 Over-arching models of feedback presented

The dominant model of feedback presented in this body of the medical education literature is one of information transfer, from the knowledgeable tutor to the naïve learner. This model is reinforced by the explicit articulation of fixed roles within the feedback process, the transfer of performance information from one party (the tutor) to another (the learner) and the use of language to reinforce this act of giving.

Sargeant et al's (2008) model of feedback and 'directed' self-assessment is linear in design, with feedback fed into the system from a source external to the learner. Ende (1983) clearly depicts students and junior doctors as the recipients of performance information from clinician; Milan et al (2006) describe clearly defined roles for 'faculty' (as the originators of feedback) and 'learners' (as recipients); and Branch & Paranjape (2002) describe the 'giving of feedback'. The work of Kluger et al (2010) and later Watling et al (2012) focus on how factors can be optimised to encourage learners to accept and integrate the feedback they receive. In their work, tutors are the creators of feedback; the model assumed is tutor-dependent and tutors are the agents of change. Schartel (2012) saw provision of feedback as a 'duty' of the tutor, although the 'sender and receiver' should work as allies. Unsurprisingly, the definition of feedback van de Ridder et al (2008) constructed after a systematic review of the medical education literature on this topic highly reinforces this information-transfer model and the receiver role as work that is designed to condense that which has gone before it cannot be expected to challenge it.

"Specific information about the comparison between a trainee's observed performance and a standard, given with the intent to improve the trainee's performance." (p.193)

A subgroup of authors are keen to recognise the limitation of a simple information transfer model of feedback but the language used during this discussion is discordant with this premise and acts to limit the autonomy of the learner and promotes adherence to this traditional model. Bing-You & Trowbridge (2009) suggested that *'the feedback dialogue has been overly centred on the role of the teacher while underemphasizing the role of the learner' (p.1330)* and that 'e*ffective feedback may require a mutual and trusting bidirectional negotiation process with give-and-take (p.1331)'.* However, his paper discussed how tutors may be failing 'at' feedback, that medical educators might not be 'providing' learners with useful feedback, and that faculty may need to 'deliver' negative feedback. Similarly, Murdoch-Eaton (2012) acknowledged the complexity of feedback and learning but focused on feedback information being 'received and understood'. Molloy (2012) recognised the limitations of 'expert' and 'passive recipient' roles and that *'didactic feedback diminishes learner agency' (p.1157).* Yet in her suggested use of regulatory focus to improve feedback, the emphasis remains on the tutor. This does not describe a change in feedback role adoption; it is a sophisticated manipulation of the naïve learner by the expert tutor.

Calls for a shift in the feedback paradigm are advocated more strongly by another subgroup of authors but, again, the language used fails to support real change. Archer (2010) recognised the limitations of a tutor-centric model of feedback and the 'diagnostic' quality of feedback in medical education. The role of the learner is promoted, through reflection-in-action and self-monitoring but this is informed by external feedback. This limits learner autonomy and describes a tutor-dependent if not tutor-centric model. Similarly, Rudland et al (2013) strongly advocated a change in feedback perspective and adoption of a student-centred model of feedback. However, their student continues to ‘receive’ feedback, is asked to ‘seek’ feedback, the tutor continues to ‘give the feedback’. Kraut et al (2015) advocated that attention should be shifted from the delivery of feedback to understanding the learner perspective. However, the feedback remains reliant on external ‘concrete sources’. Ajjawi (2012) highlighted this discord between the theory presented and language used and that through ongoing use of this language, the information transfer model of feedback is reinforced. *“Language serves as more than a vehicle for the delivery of messages; it shapes and gives rise to ideas.” (p.1018)*

The portrayal of the feedback process by Telio et al (2015) represents the closest resemblance to the desired paradigm shift. In this paper, the feedback process is reframed:

“… from one of information transmission (from supervisor to trainee) to one of negotiation and ***dialogue….*** *that involves seeking shared understanding of performance and standards, negotiating agreement on action plans, working together toward reaching the goals, and* ***co-creating*** *opportunities to use feedback in practice.” (p.612)*

The limitation of this work, however, is that how this reframing occurs, specifically what this feedback model looks like, how it is constructed and achieved, is not fully described. This work describes the qualities of a reconceptualised feedback but falls short of proposing an articulated model, as seen in more tutor-centric pieces.

2.4.6 Summary of theory

This structured literature review identified 24 papers within the medical and surgical education literature that focused on theorising feedback. In contrast to the experimental studies (the majority of which were identified within the surgical education literature), all but one of these pieces were published in the medical education literature. The majority were published in academic medical education journals.

The papers identified were subcategorised as review articles, original work and commentary pieces. The review articles and original works that focused on author opinion were largely contributed by clinicians and published in clinical journals. The original works that discussed new ideas, concepts and critique, and the commentary pieces were largely contributed by educational academics and published in academic educational journals.

Within this body of the literature, there was both representation of an information-transfer model of feedback and a call for reconceptualisation of feedback, with an emphasis on promotion of an active learner role within a learner-centred feedback model. However, a persisting confusion between assessment and feedback, adherence to tutor-dependent models of feedback and use of language that serves to limit learner autonomy present barriers in this shift towards a more dialogic and co-created feedback. Most significantly, a practical description of this reconceptualisation is not yet articulated within the existing medical education literature.

2.5 CONCLUSION

This chapter presented the findings of a structured and comprehensive systematic review of feedback literature in medical and surgical education. It illustrated the preoccupation in the surgical education literature with the exploration of the effect of feedback via experimentally designed studies. The end destination of the experimental papers, and the review and opinion-dominated articles pertaining to the theory of feedback, is most often publication in clinical journals, with a predominantly clinical readership. In contrast, articles that focus on theorising feedback, with discussion of new ideas, concepts and critique, are published in the medical education literature, with a largely academic readership. This separation of theory and investigation prevents one from informing the other.

The overall quality of the quantitative and qualitative experimental studies is poor, limiting what this body of the literature can tell us about the effect of information-transmission feedback on learner performance. Within these studies, there is no exploration of the learner role and no studies involved a dialogic feedback process.

Within the medical education literature pertaining to the theorisation of feedback, there is evidence of a reconceptualisation of feedback, from information-transfer, tutor-centric models towards a dialogic, exploratory model more in keeping with the work of Carless et al (2011) and Boud and Molloy (2013). Or rather, there is evidence of the recognition of the need for reconceptualisation. What is lacking is the understanding of what that model would look like, articulation of how it could be implemented, and what the effect on learning would be. With the purpose of exploring our understanding of an active learner role, its mechanisms and effect, a subsequent literature review relating to self-regulation of learning was undertaken.

CHAPTER 3: STRUCTURED LITERATURE REVIEW OF SELF-REGULATION

The current chapter presents the findings of a structured review of self-regulation literature in medical and surgical education. The current chapter presents the findings of a structured systematic review of self-literature in medical and surgical education. The focused aim of this literature review was to identify and analyse research articles published specifically to investigate the measurable effect of self-regulation, via an experimental design, and those that discuss the theory of self-regulation in relation to learning. In doing so, I intended to explore the currently held beliefs in relation to the effect of self-regulation within medical and surgical education, present and critique the quality of the quantifiable evidence these were based upon, and to capture the existing discourse in this literature surrounding self-regulation. This review was essential in identifying important themes relating to self-regulation but also gaps in the literature that might represent areas which new research might usefully address.

3.1 SELF-REGULATION IN MEDICAL AND SURGICAL EDUCATION

A structured literature review was conducted to identify literature relevant to the study of self-regulation in medical and surgical education. The searches were conducted via the PubMed database, a service provided by the National Centre for Biotechnology Information (NCBI) of the US National Library of Medicine. PubMed provides free access to MEDLINE®, the NLM® database of indexed citations and abstracts to medical, nursing, dental, veterinary, health care, and preclinical sciences journal articles. The PubMed journal list includes approximately 30,000 journals. These literature reviews were conducted in July 2016.

3.2 SEARCH TERMS AND RESULTS

Two separate searches using broad Medical Subject Headings (MeSH) terms and keywords (shown in quotations) were performed and the results were then combined:

- Search one: medical education AND “self-regulation”
- Search two: surgical education AND “self-regulation”

The searches were purposefully broad to maximise the potential inclusion of articles of interest in subsequent stages of the literature review. The MeSH terms “medical education” and “surgical education” were used in the searches but were combined with the keyword “self-regulation”. The MeSH term “self-regulation” was not used as this search resulted in a very small number of results (n=4). Research article titles and then abstracts were reviewed and subjected to the inclusion and exclusion criteria detailed in Box 3. Duplicate articles were removed. Figures 9-12 illustrate the stepwise process through which the summative collection of literature was attained.

Box 3: Literature search criteria

<u>**Title review: Exclusion criteria**</u>

No mention of self-regulation in the title
Non-education based self-regulation

- Clinical performance self-regulation
- Professional self-regulation

Perceptions of regulatory processes

Studies relating to self-directed learning

<u>**Abstract review: Exclusion criteria**</u>

Criteria pertaining to title reviews
Abstract (or subsequent article) unavailable in English
Abstract-only publications
Focus on assessment
Study focusing on measuring self-regulation (without relation to outcome)

Figure 9: ‘Medical education’ AND ‘self-regulation’ systematic literature review

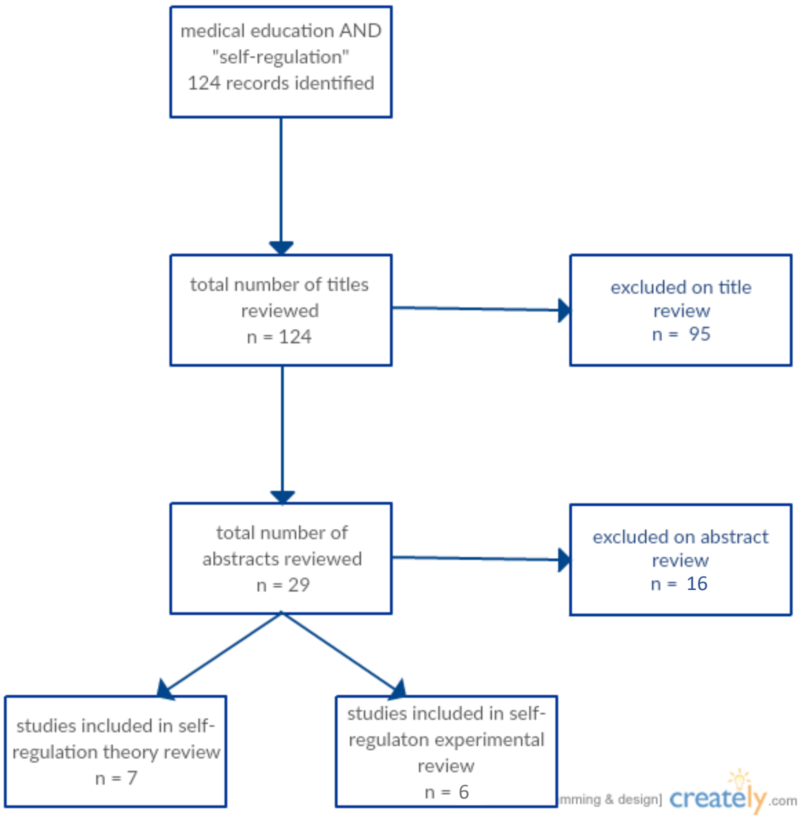


Figure 10: Surgical education AND "self-regulation" systematic literature review

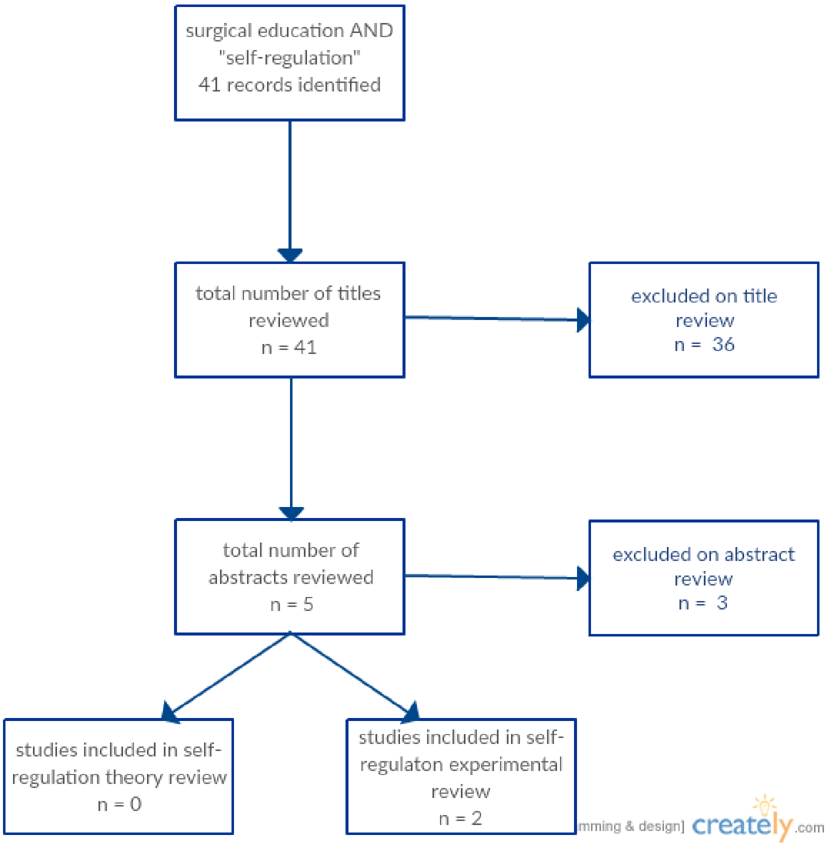


Figure 11: Self-regulation experimental studies systematic literature review

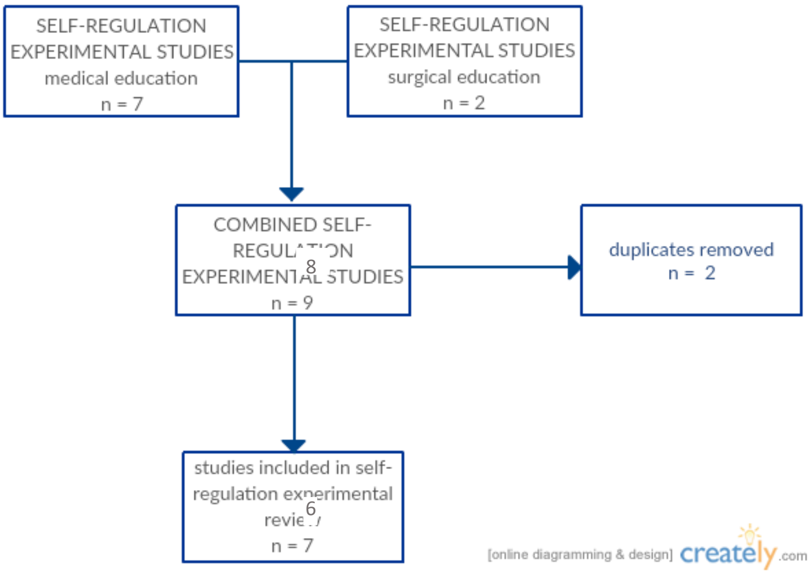
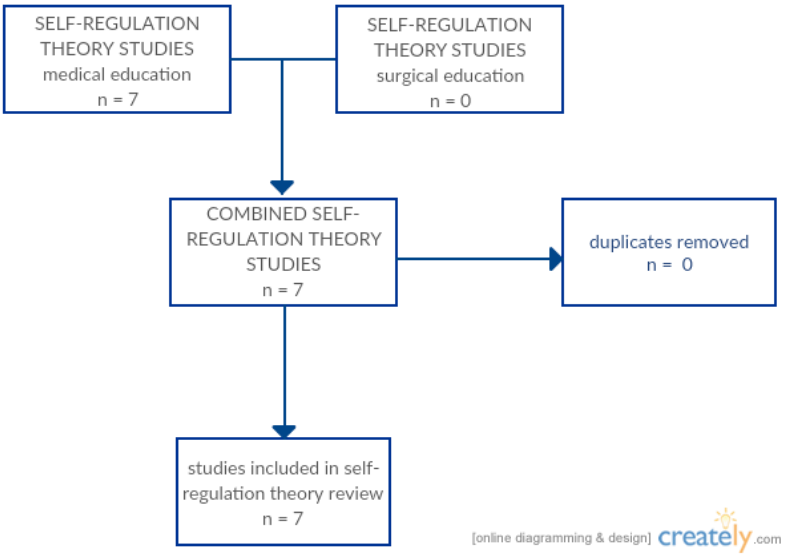


Figure 12: Self-regulation theoretical papers systematic literature review



3.3 THE INVESTIGATION OF SELF-REGULATION VIA EXPERIMENTAL STUDIES WITHIN THE MEDICAL AND SURGICAL EDUCATION LITERATURE

Via this systematic literature review, six papers were identified in which an experimental study design was used to investigate the relationship between self-regulation and academic or psychomotor task performance (Figure 11). The primary role of four authors is as a medical education academic, one is a clinician and one is a psychologist with an education interest. All of these papers were published in medical education journals. A summary of the seven papers is found below in Table 8.

Table 8: Summary of experimental papers

Investigating	Participants	Task	Performance measures	Intervention	Self-regulation measures	Design	Results
Sobral (2005) Medical students' mindset for reflective learning: a revalidation study of the reflection-in-learning scale. Adv Health Sci Educ Theory Pract, 10(4), 303-14.							
To examine whether the Reflection-in-Learning Scale (RLS) could help recognize reflective learners, and to identify whether there are relationships between RLS scores early in the medical program and later academic performance of students.	*Undergraduate* 275 medical students	*Academic* Examination performance	*Academic achievement: GPA score* *Self- confidence measure*: A visual analogue scale reflected confidence as a learner.	N/A	Score attained on a four-item version of the RLS.	Non-interventional cohort study, 2-year academic follow up RLS was administered to participants at start and at end of the third semester in Year 2.	**Recognising reflective learners:** No more than 18% of the students kept both a high RLS score and high-perceived personal efficacy on the ability to reflect at each consecutive appraisal. **GPA:** The association of RLS with sixth-semester GPA was stronger with the end-of-term measure. Change in GPA during the follow-up period showed a significant positive correlation with end-of-term RLS.
Cleary & Sandars (2011) Assessing self-regulatory processes during clinical skill performance: a pilot study. Med Teach, 33(7), e368-74.							
To evaluate the use of SRL microanalysis to assess the regulatory profiles of students who were successful and unsuccessful in a venepuncture task.	*Undergraduate* 7 medical students	*Psychomotor Bench model* Venepuncture on mannequin model	*Task performance measure:* Success of obtaining blood sample *Self-efficacy/confidence measure:* 0-100 *Self-satisfaction with task performance*: 0-100 *Self-evaluative standards:* participants asked what they used to judge their self-efficacy/satisfaction	N/A	A three-item SRL microanalytic protocol designed to assess participant cognition and self-regulation practices: *Q1 (Planning):* ‘‘What are you thinking about as you prepare to draw blood from this arm?’’ *Q2 (Goal setting):* ‘‘Do you have a goal in mind before drawing this blood sample?’’ *Q3 (Metacognitive monitoring):* ‘‘Do you think you have performed a flawless process thus far or have you made any mistakes? Tell me about them’’.	Non-interventional single-visit cohort study Single performance of task and data acquisition episode.	5 participants able to obtain blood sample at first attempt (“Successful”); 2 participants required 2 attempts before obtaining a sample (“Strugglers”). **Successful participants:** exhibited a high level of strategic thinking across all four self-regulatory processes (planning, goal setting, monitoring, and self-evaluative standards). Described process/technique focusses throughout. **Struggling participants:** displayed a non-strategic approach, being more focused on outcomes (i.e. being able to obtain a blood sample) than they were in the process of performing the venepuncture task correctly. Described outcome focusses throughout the four self-regulatory processes.

Investigating	Participants	Task	Performance measures	Intervention	Self-regulation measures	Design	Results
Brydges et al (2012) Directed self-regulated learning versus instructor-regulated learning in simulation training. Med Educ, 46(7), 648-56.							
To evaluate the effectiveness of directed self-regulated learning (DSRL) and instructor-regulated learning (IRL), on psychomotor task performance	*Postgraduate* 42 internal medicine residents (23 participated in delayed retention testing)	*Psychomotor Bench model* Simulated lumbar puncture (Lumbar Puncture Simulator II, 'easy' and 'difficult' models)	*Self-reported confidence:* via 11-point Likert scale *Task performance:* via expert-generated global rating scale (GRS) score and procedural checklist (PC) score	*Group 1 (IRL):* Group practice with instructor (4:1 ratio) (max 35min): no access to instructional video review, progress through easy-difficult simulator models based on discussion with instructor. *Group 2 (DSRL):* Self-directed practice on simulators (max 35min): open access to instructional video review, self-determined progress through easy-difficult simulator models.15min 1:1 instructor feedback and guidance after post-intervention test.	N/A	Two-visit RCT Group 1: n=22 Group 2: n=20 Simulator orientation 10-minute instructional video Pre-study confidence rating Pre-intervention performance Practice according to intervention group Post-intervention performance Post-study confidence rating Retention test performance (3 months delay)	Scores in both groups increased from pre-intervention to post- intervention on the GRS, the PC and for self-reported confidence ($p < 0.05$). IRL tended to perform better at post-intervention than DSRL group ($p=0.09$) **Self-reported confidence:** IRL group experienced a greater increase in confidence from pre-intervention to post-intervention than the DSRL group ($p=0.015$) **Retention testing:** DSRL group maintained its post-intervention performance, whereas that in the IRL group dropped significantly ($p < 0.05$). **Performance vs confidence correlations:** Positive and significant for the DSRL group, and negative and non-significant for the IRL group.
Turan & Konan (2012) Self-regulated learning strategies used in surgical clerkship and the relationship with clinical achievement. J Surg Educ, 69(2), 218-25.							
To investigate the self-regulated learning strategies used by medical students in surgical clerkship and their relationship with clinical achievement.	*Undergraduate* 273 medical students as cohorts undertaking surgical clerkship	*Academic* Examination performance	Multiple choice examination score OSCE examination score Tutor performance score.	N/A	Motivated Strategies for Learning Questionnaire (MSLQ).	Non-interventional multiple measures cohort study, 10 weeks duration	**OSCE scores**: Self-efficacy levels were positively correlated with OSCE scores, while control over learning beliefs were negatively correlated with OSCE scores. **Multiple choice examination score:** No significant relationship was defined between the MSLQ's and case-based examination scores. **Tutor performance score:** There was a weakly significant relationship between the MSLQ and tutor evaluation scores.

Investigating	Participants	Task	Performance measures	Intervention	Self-regulation measures	Design	Results
Shanks et al (2013) Are two heads better than one? Comparing dyad and self-regulated learning in simulation training. Med Educ, 47(12), 1215-22.							
To compare the relative effectiveness and efficiency of dyad versus directed self-regulated learning (DSRL) training of simulation-based lumbar puncture (LP).	*Postgraduate* 42 internal medicine residents (28 participated in delayed retention testing)	*Psychomotor Bench model* Simulated lumbar puncture (Lumbar Puncture Simulator II, 'easy' and 'difficult' models)	*Self-reported confidence and experience:* via Likert scale and number of lumbar punctures previously performed. *Task performance:* via blinded expert-generated global rating scale (GRS) score and procedural checklist (PC) score	*Group 1 (dyad):* 2:1 ratio of participant to simulator. Practiced on the simulators in pairs (max 35min) observing/practicing in equal volumes. Open access to instructional video review, self-determined progress through easy-difficult simulator models. Tutor available for consultation with frequency of consultation recorded. *Group 2 (DSRL):* 1:1 ratio of participant to simulator. Self-directed practice on simulators (max 35min). Open access to instructional video review, self-determined progress through easy-difficult simulator models. Tutor available for consultation with frequency of consultation recorded.	N/A	Two-visit RCT Group 1: n=22 Group 2: n=20 Simulator orientation 10-minute instructional video Pre-study confidence and experience rating Pre-intervention performance Practice according to intervention group Post-intervention performance Post-study confidence rating Retention test performance (2 months delay) Sample size calculation performed but study significantly over-recruited (22 vs 45 participants); no reason given.	**GRS scores:** Scores in both groups increased from pre-intervention to post-intervention to retention testing. The gain from pre- to post-intervention scores were higher in the dyad group compared to the DSRL group. **PC scores:** Scores in both groups increased from pre-intervention to post-intervention testing. **Retention testing:** For the GRS and PC scores, both groups improved from pre-intervention to post-intervention and showed a small drop in performance from post-intervention to retention test. There were no significant inter-group differences (p=0.58). **Self-reported confidence:** Scores in both groups increased from pre-intervention to post- intervention

Investigating	Participants	Task	Performance measures	Intervention	Self-regulation measures	Design	Results
Artino et al (2014) Exploring clinical reasoning in novices: a self-regulated learning microanalytic assessment approach. Med Educ, 48(3), 280-91.							
To examine the regulatory processes during a diagnostic reasoning task and correlate the quality of strategic thinking with academic performance in standardised examinations	*Undergraduate* 71 medical students (“novice learners”)	*Academic* Diagnostic reasoning task (Synthesis of differential diagnoses in relation to hypothetical clinical case)	*Examination scores:* GPA, USMLE and NBME	N/A	A three-item SRL microanalytic protocol designed to examine medical students’ intra-task regulatory processes (goal setting, strategic planning and metacognitive monitoring): *Q1 (Immediately after reading the case):* **‘**Do you have a goal (or goals) in mind as you prepare to do this activity? If yes, please explain.’ *Q2 (Immediately after Q1):* ‘What do you think you need to do to perform well on this activity?’ *Q3 (Asked immediately after first differential diagnosis recorded):* ‘As you have been going through this process, what has been the primary thing you have been thinking about or focusing on?	Non-interventional multiple measures cohort study, 1-year duration. Single performance of task and data acquisition episode. Multiple measures of academic performance (GPA, USMLE score, NBME score)	**Forethought phase (Q1&2):** Although 32% of participants did provide a strategic goal, approximately 50% of participants either conveyed goals that focused on the outcome of getting the correct diagnosis (18%) or did not report any type of goal (31%). Strategic planning was statistically significantly correlated with goal setting ($p < 0.01$) **Performance phase (Q3):** Unlike the pattern of results observed for the forethought phase processes, 90% of students reported that they were focused on task-specific processes while they completed the diagnostic reasoning task. **Exam performance:** Strategic planning explained significant variance in second-year GPA ($p < 0.01$), USMLE score ($p<0.05$) and NBME score ($p<0.05$); with moderate effect size. Students who were focused on several task-specific processes as they approached the diagnostic reasoning task achieved better results on both short- and longer-term performance outcomes.

3.3.1 Broad study themes and aims

Within this group of six papers, there are two different broad themes: studies that explore the relationship between self-regulation and psychomotor task performance; and those that explore the relationship between self-regulation and academic performance. A summary of this breakdown is shown in Table 9.

Table 9: Breakdown by study theme

Exploration of self-regulation + psychomotor task performance	Exploration of self-regulation + academic performance
Cleary (2011)	Sobral (2005)
Brydges (2012)	Turan (2012)
Shanks (2013)	Artino (2014)

Cleary and Sandars (2011), Brydges et al (2012) and Shanks et al (2013) all explored the relationship between learner self-regulation and psychomotor task performance. Cleary and Sandars (2011) attempted to capture and measure the real-time self-regulatory processes employed during venepuncture, via a microanalytical protocol, and correlate these findings with success or failure of task completion. Brydges et al (2012) and Shanks et al (2013) – both of whom were a co-author on the other paper – utilised similar methods in their interventional studies. Both of these papers examined how variation in learning environment, with different levels of learner autonomy, affected the quality of lumbar puncture performance.

Sobral (2005), Turan and Konan (2012) and Artino et al (2014) focused on the relationship between learner self-regulatory practices and academic performance; quantifying this via a combination of measures external to the study itself. These included end of undergraduate block tests (written and OSCE examinations, tutor-awarded grades) and end of year examinations (Grade Point Average, United States Medical Licensing Examination and National Board of Medical Examiners) (Table 8).

An important observation in relation to this area of the literature is that studies were either of an interventional design and measured psychomotor task performance but not the degree of participant self-regulation (Brydges et al, 2012; Shanks et al, 2013), or they were of a non-interventional design but measured psychomotor/academic performance and the degree of participant self-regulation (Cleary & Sandars, 2011; Sobral, 2005; Turan & Konan, 2012; Artino et al, 2014). No study methodology involved the implementation of an intervention, the measurement of self-regulatory processes and the success or quality of psychomotor/academic performance.

3.3.2 Study design

Two of the studies that explored the relationship between the degree of self-regulation of learning environment and psychomotor task performance (Brydges et al, 2012; Shanks et al, 2013) employed two-visit randomised control design, with the intervention enacted and most of the data (single pre- and single post-intervention performance) collected during the first visit. The second visit was brief and allowed for delayed retention testing, at two and three months respectively. The third of these studies (Cleary & Sandars, 2011) employed a single-visit cohort study design, with performance and self-regulation data pertaining to only 10 venepunctures comprising the total study data.

Cleary and Sandars (2011) recruited only seven undergraduate participants to their self-regulation microanalysis pilot study. Participant numbers recruited to the other two psychomotor task studies were larger; recruiting 42 (Brydges et al, 2012) and 45 (Shanks et al, 2013) postgraduate participants. Interestingly, Shanks et al (2013) included a sample size calculation in their study method but over-recruited by 100% (45 vs 22 participants), overpowering the study and increasing the risk of type I error (a false positive), without explanation. This limits the quality of the design of this study and the subsequent conclusions that can be drawn.

The studies which explored the relationship between learner self-regulation and academic performance (Sobral, 2005; Turan & Konan, 2012; Artino et al, 2014) were all cohort studies involving undergraduate participants. These studies tended to be of longer duration (two years, 10 weeks and 1 year respectively), which is necessary given their use of summative academic results as measures of academic performance. These studies recruited larger cohorts than their psychomotor task study counterparts; with between 71 and 275 participants recruited, without evidence of a sample size calculation.

3.3.3 Measures of self-regulation

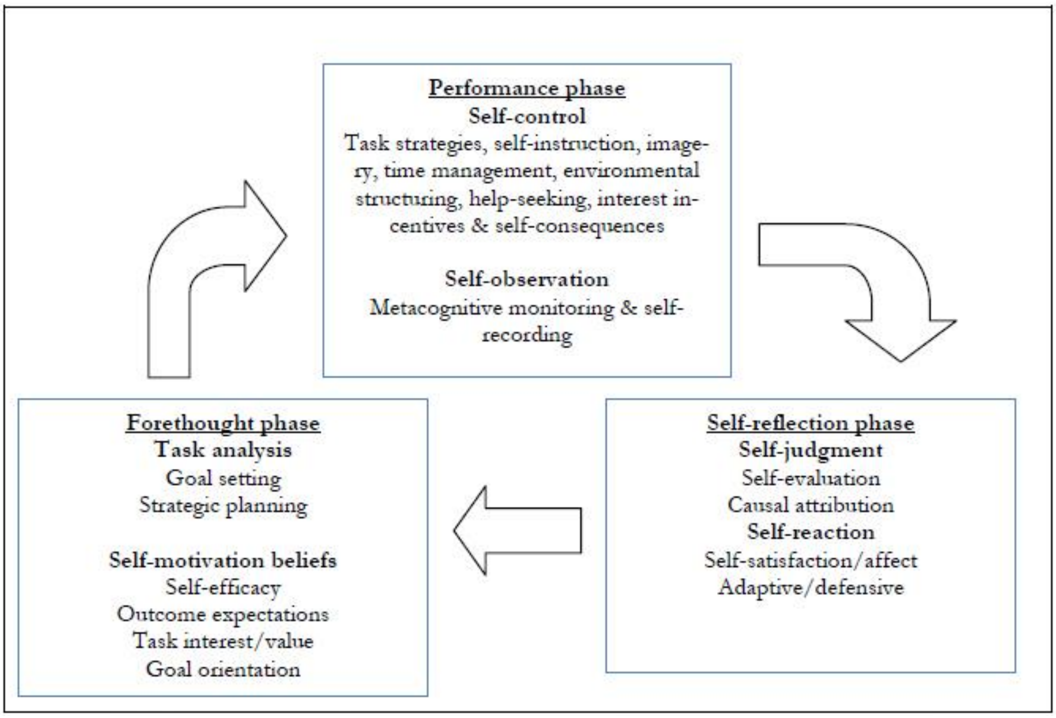
Four of the studies identified by this literature review (Sobral, 2005; Cleary & Sandars, 2011; Turan & Konan, 2012; Artino et al, 2014) featured a measure of self-regulation. Cleary & Sandars (2011) and Artino et al (2014) developed and utilised a microanalytic method, using questions during participant task performance to detect the presence and type of strategic thinking being utilised. Questions were asked at specific strategic points during the task performance, and analysis of the answers given allowed the authors to categorise responses, and participants, as process, outcome and more general thinkers. This tool was piloted by Cleary and Sandars (2011) in relation to a venepuncture task and then utilised by Artino et al (2014) in a larger study involving a diagnostic reasoning task; proving applicable to both psychomotor and academic tasks. This microanalytic process is designed to detect and explore real-time intra-task self-regulation of performance.

The two studies that explored long-term regulation of learning (Sobral, 2005; Turan & Konan, 2012) utilised more suitable tools for that task; the Motivated Strategies for Learning Questionnaire (MSLQ) and a modified Reflection-in-Learning Scale (RLS). These validated tools gauge the learner's perceptions towards more long-term factors related to learning, such as motivation, goal orientation, beliefs in relation to control of learning, and strategies employed to promote learning. Analysis of these results allows conclusions to be drawn in relation to the self-regulatory behaviours in relation to continuing and longstanding learning. Therefore, these scales take a more overarching and extra-task measure of self-regulation of learning.

3.3.4 The definition of self-regulation

As discussed in chapter one, Zimmerman's (2000) definition of self-regulation as *'self-generated thoughts, feelings and actions that are planned and cyclically adapted to the attainment of personal goals'* (p. 13) is accepted as our definition. Figure 13 illustrates the three key processes involved in self-regulation: forethought (with utilisation of process goals); performance (with active self-monitoring); and reflection (consideration of process to inform ongoing self-regulation cycles).

Figure 13: Cycle of self-regulation (Reproduced from Zimmerman and Moylan, 2009)



Within this body of literature, the term "self-regulation" is used to convey three broadly different functions: self-regulation of task performance as a skill, self-regulation of learning and self-regulation of learning materials and environment.

Cleary & Sandars (2011) and Artino et al (2014), in their microanalyses of self-regulatory practices during task performance, discuss self-regulation as an intra-task practice. They describe a real-time planning, self-monitoring, and self-reflection process that occurs before, during and after task performance. Although the authors specifically use the term "self-regulated learning", both studies focus on the planning, goal setting and metacognitive elements of the task 'planning' and 'performance' phases of the self-regulation cycle, rather than the 'reflective' and feedforward elements associated with the reflection phase. These studies focus on the recording, categorisation and analysis of the participant's self-regulatory practices during psychomotor and academic task completion. Artino et al (2014) cite self-regulation as 'skill', further identifying and supporting this task performance orientated definition.

The work of Sobral (2005) and Turan and Konan (2012) was concerned with the correlation between more long-term self-regulation of participant's learning practices and academic achievement. Turan and Konan (2012) mirrored the previous description of self-regulation as a cyclical process, with the construction of goals, and monitoring and control of progress but the focus of this process is learning itself, not specific task completion. This focus on learning is shared by Sobral (2005), who promotes the use of reflection *'as a process of managing and adjusting the progress of learning while it is taking place' (p.305)*. The tools they used to capture self-regulatory practices were primed to collect data regarding more long-term ideas, attitudes and practices that related to learning, not specific task completion.

The central focus on 'directed self-regulated learning' (DSRL), illustrates that Brydges et al (2012) and Shanks et al (2013) investigated a type of learning model. In these studies, self-regulation is not regarded as a skill employed during task completion, or a facet of learning that can be promoted, but a type of learning style and environment, facilitated by materials, resources and tutors, in which greater independence of the learner is promoted. In these interventional studies, the effect of learning environment on participant task performance was investigated. As the focus shifts from learner to learning environment, the language used changes as well. Brydges et al (2012) emphasised the vitality of a skilful and knowledgeable tutor in designing directed self-regulated learning conditions. Shanks et al (2013) cites the efficiency of dyad practice as one of the major advantages of this model. This efficiency applies to materials, equipment and instructors rather than learners, who took similar lengths of time regardless of learning model. This language serves to promote a shift from a learner-centric to tutor-centric discussion, with the emphasis on the role of the tutor in orchestrating an appropriate learning environment.

3.3.5 Results: the effects of self-regulation

3.3.5.1 Promoted self-regulation of learning environment

Collectively, the studies of Brydges et al (2012) and Shanks et al (2013) compared the effects of instructor-regulated learning (IRL), directed self-regulated learning (DSRL) and dyad directed self-regulated learning (DDSRL) on psychomotor task performance. Performance was measured by two subjective quantitative scores (expert generated Global Rating Scale and Procedural Checklist scores). The studies suggest that all three models of learning statistically improve performance between pre- and post-intervention testing. Brydges et al (2012) demonstrated an inter-group difference between IRL and DSRL at delayed retention testing, three months post study, with the DSRL group maintaining previous performance more successfully. Shanks et al (2013) showed no such inter-group difference between DRSL and DDSRL at retention testing at two months. The authors cited the limitations of concurrent feedback as explanation of the inferior performance of the IRL group and the greater facilitation of self-monitoring for the improvement of skill retention displayed by the DSRL group.

3.3.5.2 Self-regulation of learning

The studies of Sobral (2005) and Turan and Konan (2012) correlated the degree of learner self-regulation with academic achievement. Sobral (2005) demonstrated that a low proportion of learners (18%) remained consistently highly self-regulatory in relation to their learning during the 2-year period. However, this work supported the hypothesis that higher levels of self-regulation (as measured by the Reflection-in-Learning Scale) was associated with a higher Grade Point Average. Whilst Turan and Konan (2012) failed to show any significant correlation between self-regulation (as measured by the MSLQ: Motivated Strategies for Learning Questionnaire) and end-of-clerkship written examination score, the study did identify a positive correlation between the self-efficacy element of the MSLQ score and OSCE examination performance, and total MSLQ score and tutor evaluation score.

3.3.5.3 Self-regulation of task performance

The prevalence and quality of intra-task self-regulation skills were investigated by Cleary and Sandars (2011) and Artino et al (2014). In brief, these studies support the assertion that participants who command a greater degree of process-focused self-regulation during task performance are both able to perform tasks more successfully and efficiently (Cleary & Sandars, 2011) and also perform better in long-term academic measures (Artino et al, 2014).

Cleary and Sandars (2011) found that participants that were successful in obtaining a blood sample in a simulated venepuncture exhibited higher levels of strategic thinking during task completion as compared to those who were unsuccessful. Successful participants described process and technique focuses during task completion. This is in contract to unsuccessful participants, who tended to display a non-strategic approach, being focused on outcomes (such as being able to obtain a blood sample) rather than the process of performing the venepuncture task correctly.

Artino et al (2014) utilised the microanalysis method to explore the intra-task completion self-regulatory function of undergraduate participants in relation to a diagnostic reasoning task. However, rather than relate this to competency of this task, the author compared it to academic achievement in several end of year summative examinations. This microanalysis technique revealed that participants were more likely to report a focus on task specific processes during the performance phase of the task (90%) as compared to the forethought and planning phase (32%). Statistical analysis accounted for participants' previous academic record. The degree of strategic thinking was a significant factor in explaining variation in participant second-year GPA ($p<0.01$), USMLE score ($p<0.05$) and NBME score ($p<0.05$), with moderate effect size. Participants who reported focus on several task-specific processes achieved better academic results. These results suggest that a higher degree of self-regulatory skill is associated with better academic performance, although it is not possible to assert that it is necessarily a causal relationship.

3.3.6 The limitations of these experimental studies

There are several quality issues affecting interpretation of these studies. There is large variation in sample size (n = 7 – 275) with no reference to sample size calculation in four of the five quantitative studies and disregard for the sample size calculation in the remaining study (Shanks et al, 2013). The studies that adopted a non-interventional cohort deign do not allow for characterisation beyond simple association of any of the relationships suggested between self-regulation and performance. The conclusions based on the results of the retention testing in the interventional studies was compromised by high rates of participant drop-out (Brydges et al, 2012 and Shanks et al, 2013; 46% and 34% respectively).

Furthermore, there is a relative paucity of studies and data to examine. All six papers identified further research as being necessary to test the validity of the results identified or to further the understanding of the relationships discussed. This last point introduces the main limitation of this group of studies: the sole focus of this area of the literature is the quantification of effects of self-regulation and not the understanding of these effects and exploration of the self-regulation process (i.e. the why and how). This relates to study design; no interventional study included a measurement of participant self-regulation; and no study included a qualitative element, which may might explore the effects of self-regulation in a way quantitative data could not.

3.4 DISCUSSION OF THE THEORY OF SELF-REGULATION IN THE MEDICAL AND SURGICAL EDUCATION LITERATURE

Via this systematic literature review, seven papers with six different first named authors were identified in which a major focus was the discussion of the theory of self-regulation (Figure 12). Six of these papers were published in medical education journals and the seventh in an academic medical journal. The primary role of three authors is as a medical education academic, two are clinicians and one is a psychologist with an education interest. Two of the authors contributed work as second author to another piece of work included in the review (Butler collaborated with Brydges (2012); Brydges collaborated with Butler (2013)). In similarity with the trend observed in the structured medical and surgical education literature concerning the theory of feedback, none of these articles stemmed from the surgical education literature. Thus, this discussion and analysis of the body of work pertaining to the theory of self-regulation is exclusively a representation of work published within the medical education literature.

3.4.1 The design and thematic analysis of published work

3.4.1.1 Original work

Of the seven papers included in this targeted literature review, two can be considered as original works (Sandars, 2011; Brydges & Butler, 2012). Summary details of this group of articles are given in Table 10. Sandars' work (2011) sought to provide a comprehensive theory-to-practice discussion of self-regulation in the context of medical education. This 12-page piece, which promotes the encouragement of self-regulation to improve academic and clinical achievement, is dominated by three themes: concepts, summary and recommendations.

Table 10: Original work

Author	***Year***	***Journal***	***Title***
Sandars	*2011*	*Med Teach*	Self-regulation theory: applications to medical education (AMEE Guide no. 58)
Brydges	*2012*	*Med Educ*	A reflective analysis of medical education research on self-regulation in learning and practice

The author explains in detail the concept of self-regulation (citing Zimmerman's definition) and the three phases (forethought, performance and self-reflection) of the cyclical process with the purpose of educating the reader. Supporting literature is summarised. In particular, the role and value of process (as opposed to outcome) goals in improving task performance are advocated, with evidence cited from both medical education and sports literature; and the author discusses the role of the tutor, the role of learning development courses and the role of feedback in promoting the self-regulatory process as represented in the self-regulation literature.

Of specific interest, the author highlights the principle that effective self-regulation relies upon learners generating accurate internal feedback or this being provided by tutors. In this way, the feedback is not designed to facilitate but to replace internal feedback and performance judgements; external feedback is referred to as a 'reality check'. An information transmission model of feedback persists, as designated by the 'giving' and 'receiving' of feedback from tutor to learner.

The recommendations contained within this work present a spectrum in relation to scope. Very specific recommendations are made with regards the role of the tutor during teaching, with detailed description of the tutor-led questions that can be employed to encourage learner self-regulation during academic and clinical activities. Broader recommendations are made in relation to the importance of the role of the tutor, integration of self-regulated learning into curricula, and regarding self-regulation as a skill, placing importance on facilitating learner understanding and ability in this skill via specific learning development courses. Finally, the conclusion of this work could be summarised as broad recommendations, with further research relating the process of self-regulation, the cognitive and emotional aspects of professionalism, gender and cultural factors advocated.

In addition to the three themes discussed above, the work of Brydges and Butler (2012) contained the generation of 'new ideas'. This comprehensive work comprised of an analysis of existing medical education self-regulation literature as applied to their proposed model of self-regulated learning (Figure 14).

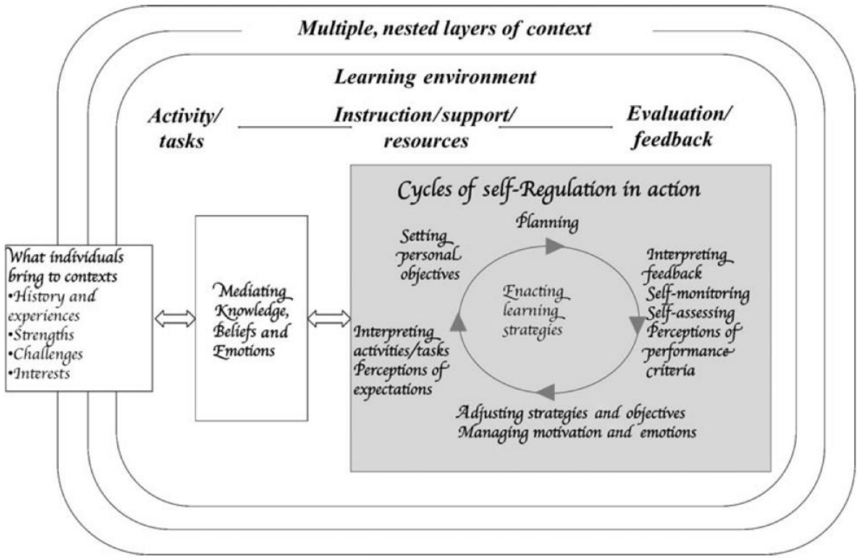


Figure 14: Brydges & Butler's model of self-regulation within medical education (Reproduced from Brydges & Butler, 2012)

A broad focus was adopted in relation to the concept of self-regulation. This work concentrated on the framing of the self-regulation process; focusing more on the factors influencing self-regulation (socio-cultural perspectives and the perceptions of the individual) than on the mechanical process of self-regulation itself. Although the 'cycles of self-reflection in action' lies at the heart of the model of self-regulation, there is little direct analysis of the more practical elements of the process. Reference to reflection-in-action is linked to self-assessment, which serves to confuse these two processes.

The recommendations contained within this work relate to the support of self-regulated learning (in a wider educational context), rather than specifically the real-time act of facilitating self-regulation. The author suggests avoiding two specific assumptions in the future application of self-regulation theory within medical education: that designing an activity for independent completion by the learner is sufficient for the development of self-regulation;

and that self-regulation is a process that is completely independent of the tutor. This promotion of the role of the tutor, with support for co-regulated practice, echoes Sandars' (2011) previous assertions.

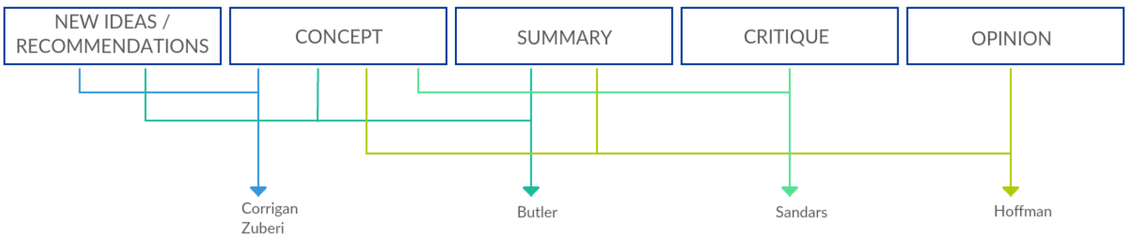
3.4.1.2 Commentary pieces

Table 11: Commentary articles

Author	***Year***	***Journal***	***Title***
Corrigan	*2012*	*Med Educ*	Self-regulated learning in medical education: the next steps
Sandars	*2012*	*Med Educ*	Future direction for research in self-regulated learning in medical education
Zuberi	*2012*	*Med Educ*	Layers within layers ... self-regulation in a complex learning environment
Butler	*2013*	*Med Educ*	Learning in the health professions: what does self-regulation have to do with it?
Hoffman	*2015*	*Acad Med*	Using self-determination theory to improve residency training: learning to make omelettes without breaking eggs

Five of the seven articles included in this targeted literature review are commentary pieces, published specifically as such and in response to the work of others (Corrigan, 2012; Sandars, 2012; Zuberi, 2012; Butler & Brydges, 2013; Hoffman, 2015). Summary details of this group of articles are given in Table 11. The dominent themes represented in these papers are similar to those in 'original work' with the addition of opinion and critique. An illustration of the combined presence of these themes is given in Figure 15. These short pieces offer the opportunity for the author to offer targetted comments and ideas, without the burden of a comprehensive summary necessitated by a review article or presentation of original work.

Figure 15: Commentary articles: the combination of themes by author



Sandars (2012) reflects on the work of Brydges and Butler (2012) and concludes that there is little practical guidance offered on how the information presented in this reflective analysis can inform teaching practices. The author went on to propose, via reference to his own collaborative work (Cleary & Sandars, 2011), that the technique of microanalysis can be used as a tool during teaching to clarify and further instruct the self-regulatory processes employed by the learner. Of interest here is the lack of concordance between Sandars' narrow focus of intra-task regulation, and Brydges and Butlers' broader focus on self-regulation of learning. Zuberi (2012) adopted a different focus when considering the work of Brydges and Butler (2012) and reflected that successful navigation of the complex, layered contexts in which self-regulation exists and can be promoted, requires the recognition and remediation of unconscious faculty cultural biases that might hinder student support.

Corrigan (2012) added to the developing published conversation regarding self-regulation by supporting an extended use of microanalytic exploration of self-regulation but postulating that alone, this was not enough. In explicit support of the work of Sandars and Cleary (2011) and Brydges and Butler (2012), he advocated bold curricular developments that actively support students' self-regulation of learning. The specifics of the form of this support and the level in which it could be integrated into curricula is not discussed.

Butler, in collaboration again with her previous co-author Brydges (Butler & Brydges, 2012), distil and reassert their earlier conclusions in their briefer commentary piece. Prompted possibly by Sandars' reflections (sandars, 2011) regarding the lack of practical instructions stemming from their review of the literature, Butler and Brydges (2012) stated that greater opportunity for learner development of self-regulation is needed; and that there are educational options beyond the dichotomy of didactic tutor instruction and learner self-discovery. They champion the role of the tutor in encouraging self-regulation, with self-regulated learning not just a pseudonym for unsupervised learning.

Hoffman (2015) described a potential basis for this facilitative tutor-learner relationship via Vygotsky's psychological theory of 'scaffolding'. In this model, the tutor provides layered external support, allowing the learner to construct understanding and skill. As this skill and understanding becomes more fully formed and independent, the support is removed (referred to as 'fading'). The argument is oriented towards the inherent tension between learner autonomy and supervision in clinical training, which is interesting but less relevant to self-regulation theory. However, Hoffman's description of the 'scaffolding' concept articulates concisely the next step in understanding what tutor-facilitated self-regulation would look like.

3.4.2 The confusing terminology of self-regulation

Self-regulation is such a broad concept that, firstly, various terms are used in the literature when discussing it and, secondly, there is contradiction in application of these terms. Whilst more than one self-regulation related term is required to define and discuss the different facets of this concept and process; the inconsistency with which they're applied serves to confuse discourse.

Sandars and Cleary (2011), adopt a tight, process-driven definition of self-regulation but applies it to both task performance, as a process, and long-term learning episodes, as an educational concept. The task-focused view of self-regulation is congruent with the microanalysis method of facilitation recommended by one author in his later commentary piece (Sandars, 2012) but is incongruent with its application to the self-regulation of learning over a longer period.

This wider view, one of self-regulation of learning, is the interpretation Brydges and Butler adopt in their co-authored reflective analysis and commentary pieces (Brydges & Butler, 2012; Butler & Brydges, 2013). In both publications, consideration of the environment in which self-reflection occurs, and is encouraged, is of primary interest. There is mention of 'self-regulation in action' but mostly the discussion relates to self-regulation of learning; a process through which learners identify missing knowledge and engage with learning materials in a self-determined way to fill them. Within the literature, this self-regulation of learning, can be referred to as self-directed learning (Shanks et al, 2013). The apparent interchangeability of these terms is clarified well by Corrigan (2012), who suggests that the former is an action of the learner, whereas the latter is a feature of the environment.

3.4.3 Clarifying the terminology of self-regulation

The difficulty is that the term self-regulation confers both a real-time, task-specific process (Cleary & Sandars, 2011; Artino et al, 2014) and a cyclical learning method adopted and applied over longer learning periods (Sobral, 2005; Turan & Konan, 2012). The conundrum of this language barrier can be navigated by considering 'grain size', as first explained by Winnie and Perry (2000) and recently proposed by Brydges and Butler (2012). Grain size refers to the level of detail of learning behaviour being considered. In this fashion, the highly specific formation of process goals, self-monitoring of performance and subsequent self-reflection in relation to an explicit task could be considered fine grain analysis. Microanalysis techniques capture this level of self-regulatory actions well (Sanders, 2012). In contrast, the process in which a learner selects learning materials and self-determines access and use of these materials to achieve larger more complex learning achievements is a large grain activity. The attitudes that underpin these behaviours may be measured by self-regulation measures such as the RLS and the MSLQ but the behaviours themselves may be better captured and understood via qualitative methods of investigation.

The two levels of self-regulation described are not distinguished by time or duration but the scope of their effect (narrow versus broad). The two levels are not independent of one another, with interaction between the two affecting long-term retention, knowledge and skill transfer, and adaptive expertise (Butler & Brydges, 2013). However, some separation of fine and large grain may provide some clarity in future discussion within the literature.

3.5 SUMMARY OF THE SELF-REGULATION LITERATURE

This structured literature review identified only six papers within the medical and surgical education literature that employed an experimental design to investigate the effect of self-regulation on learning. Similarly, it identified only seven papers within the medical education literature that focused on the discussion of theoretical application of self-regulation to learning.

Studies with an experimental design would suggest that learners who engage in directed self-regulated learning perform better than those engaged in instructor regulated learning at delayed testing; that learners who display higher self-regulatory tendencies perform better in end of year examinations; and those who display higher levels of intra-performance self-regulation behaviours are more successful when performing psychomotor skills. However, the number of studies available was small and the overall quality was poor. None of the interventional studies included a measurement of participant engagement with self-regulation. Finally, the focus of this subsection of the literature was only the attempted quantification of the effects of self-regulation, not the understanding of these effects and exploration of the self-regulation process.

The subsection of the literature focused on the discussion of the theory of self-regulation and its application to medical education provides debate of concepts, recommendations and proposes a model of self-regulation in practice. However, the lack of clarity of terminology limits the discussion. Furthermore, the proposed model relates to self-regulation of learning rather than of task and the lack of practical framing limits its ready application to educational practice.

CHAPTER 4: IDENTIFYING RESEARCH PURPOSE & DESIGN CONSIDERATIONS

4.1 IDENTIFYING GAPS IN THE FEEDBACK AND SELF-REGULATION LITERATURE

The medical education literature concerned with the theory of feedback recognises the potential benefit of a shift from a linear information transfer to a dialogic, learner-centric model, as discussed in the wider education literature. However, the language used within the medical education literature serves only to sustain the dominant tutor-dependent model, in which feedback is given and received. Furthermore, this change in perspective is not echoed in the literature concerned with the experimental investigation of feedback. Research on the effectiveness of feedback in medical and surgical education continues to be dominated with the quantification of the effects of instructor-dependent modes of feedback with little assessment of skill retention and transfer.

Work within the medical education literature concerned with the experimental investigation of self-regulation has attempted to explore the relationship between self-regulatory behaviour and academic and psychomotor task performance. However, these studies have not been able to promote and measure learner self-regulation whilst quantifying the effect on performance. Neither the self-regulatory nor feedback experimental studies have explored the effect of promoted self-regulation on learner experience. The self-regulation theory-related literature has identified the potentially powerful effects of promotion of self-regulatory behaviour. However, it has not yet explicitly linked this concept to promotion via a self-regulation focused feedback model.

4.2 AN INTEGRATED MODEL OF DIALOGIC FEEDBACK WITH PROMOTION OF SELF-REGULATION

In this study, I aimed to investigate the effect of dialogic feedback with encouraged self-regulation of learning on psychomotor task performance, skill retention and the learner experience of feedback.

Based upon consideration and analysis of the feedback and self-regulation literature, I proposed the following characterisation of two different feedback models for comparison:

traditional linear information transfer feedback and dialogic feedback with encouraged self-regulation (Box 4). Practical differentiation of these models was essential for designing further investigation.

Box 4: Feedback models and their characteristics

Features of information transfer feedback	**Features of dialogic feedback with self-regulation**
Tutor driven	Tutor facilitated
Content largely given by tutor	Content largely drawn from learner
Learner role is passive	Learner role is active
Discussion is directive	Discussion is exploratory
Focus is on outcome of actions	Focus is on process
Goal-orientated behaviour not promoted	Process goal-orientated behaviour promoted

4.3 RESEARCH QUESTIONS

I proposed two research questions for the investigation of the effect of dialogic feedback with encouraged self-regulation:

1. What is the effect of an integrated model of dialogic feedback with encouraged self-regulation versus an information transfer model of feedback on psychomotor task performance and longevity of skill retention?
2. What is the effect of the integrated model of dialogic feedback on learner's experience and understanding of feedback?

4.4 DESIGN CONSIDERATIONS

4.4.1 Adopting a mixed methodology

Addressing these research questions requires both a quantitative a qualitative approach and, therefore, this study lends itself to a mixed methodology.

The investigation of the effect of one feedback model versus another (information transfer and dialogic) on the performance of a psychomotor task suggests an examination of cause and effect and the generation of empirical data to create new knowledge, not contained within the current body of literature. With respect to this facet of the presented research, a positivist paradigm was adopted, and a deductive approach was utilised (Tavakol & Sandars, 2014). Subsequently, it was decided that the study design should include a quantitative element, with the purpose of measuring the effect of these feedback models.

The design of this quantitative arm of the study may have utilised learning episodes naturally occurring in medical education, such as the performance of clinical or surgical procedures, but this poses significant practical obstacles in relation to standardisation of conditions and accuracy of data collection. Such practical difficulties might serve to diminish the objectivity of the data collected and undermine its validity. Therefore, it was decided that this cause and effect relationship would be investigated within a simulated environment, where standardisation would be promoted, and conditions and the psychomotor tasks being measured could be controlled closely.

However, the investigation of the effect of a feedback model on a learner's experience and understanding of feedback requires a different approach. This question is exploratory in nature and cannot be answered using quantitative methods. Instead, this study must include a qualitative element, in which the naturalistic paradigm is adopted, and an inductive approach implemented (Tavakol & Sandars, 2014).

Furthermore, the mixed methods adopted should not be seen as antagonistic but instead synergistic. It is hoped that the quantified effect of these divergent feedback models might be better understood by their interpretation in the context of greater understanding of the learner experience. In reciprocation, discussion of the ideas expressed by learners regarding their experience and understanding of feedback might better inform, particularly in relation to the experimental dialogic feedback model, how this feedback model might be created.

4.4.2 Quantitative design considerations

The quantitative investigation of the effect of feedback on psychomotor task performance will be conducted via a statistically robust experimental study, featuring randomised control design, multiple and repeated measures of performance. More specific deign decisions (psychomotor task, visit protocol and measures of performance) will be shaped by subsequent pilot investigation as discussed in later sections of this thesis (Chapter 5: Pilot study one, Chapter 6: Pilot study two; and Chapter seven: Full study methods).

4.4.3 Qualitative design considerations

The effects of feedback model on learner experience and feedback literacy will be addressed in the qualitative element of the study via participant semi-structured interview, as I believe this method will best align with the research purpose (Tavakol & Sandars, 2014). Conducting interviews will allow participants to describe their thoughts, ideas and perspectives in their own words, promoting richness of data, and a semi-structured design will allow for direction of this discussion towards areas of research interest. The specific interview questions will be based upon identified gaps in the current feedback literature and realised via an iterative process throughout pilot study conduction (section 9.1.1).

Alternative options for qualitative design include conducting focus groups or unstructured in-depth interviews (Tavakol & Sandars, 2014). In relation to the former, data collected via focus groups is socially constructed and whilst this lends strength when the sharing of views and deliberation is valuable, the ideas expressed as part of focus groups may not represent all participants views. I am concerned that this method of data collection would lead to a narrowed discourse and loss of potentially valuable expression. Reciprocally, it is possible that completely unstructured interviews might fail to produce discussion of key, identified areas of specific interest, making them less suitable for this research purpose.

4.4.4 Population and sampling

As discussed within the literature review, most existing quantitative studies investigating the effect of feedback have drawn samples from the medical student population. Other potential populations included Foundation doctors (doctors within the first two years post-graduation) , surgical or medical trainees (a minimum of two years post-graduation but a larger range in terms of experience and maturity depending on length of training and volume of pre-training employment), or even fully qualified clinicians.

For promotion of generalisability, minimisation of bias, and practical convenience, it was decided that a convenience sample would be drawn from undergraduate medical students. Within the context of the study's significant quantitative element, reducing variation within the population sampled would promote quality within the study and this group contains the least variation in relation to experience of psychomotor task learning. By virtue of their predictable learning programme, participants were unlikely to be exposed to significant variation in volume of psychomotor task performance or learning during the study, limiting potential contamination of results. As the study was conducted at Dundee University & Medical School, even potentially large numbers of undergraduate medical students would be convenient to sample.

In order to further quantify and potential baseline differences between participants and acknowledge any potential confounding factors, pre-study demographic data will be collected. This will include age, gender, dominant hand (as many psychomotor tasks are asymmetrical and therefore, may favour one hand dominance over another), self-rated confidence level, self-rated experience level, description of experience, and participant confidence in tutor (researcher).

Whilst recruiting medical students to a medical education study confers a degree of authenticity and validity of the application of findings within this population, it should be acknowledged that the generalisability of the findings might be limited to other novice learners, with caution applied in relation future results to more advanced learners. In order to mitigate this limiting factor to a degree, and to further reduce variation within the study population, it was decided that the study sample would be drawn from final stage students only, in their fourth and fifth years of study of a five-year programme. Such that this sample was not contaminated nor limited, samples for pilot stages of investigation were drawn from more junior years of study (second year for pilot study one; third year for pilot study two).

Furthermore, generalisability may be limited by the inherent bias of voluntary recruitment to the study, as participation may appeal to the more motivated students, or those with an early surgical career aspirations, or even latent psychomotor ability. However, implications regarding generalisability withstanding, these limitations would affect both study groups equally, and it would be unlikely to be accountable for any inter-group differences observed.

4.4.5 Reflexivity

Reflexivity refers to the principle of acknowledging the context of research findings, and in particular to acknowledging and attending to the effect that a researcher will exert during the research process. It is imperative that I acknowledge my own position and perspectives, particularly given my ‘insider researcher’ role in this study. My involvement in this study will affect every stage of its development, from premise and design, how the study is conducted and how participants may behave, to analysis and understanding extrapolated from study results.

From an early stage in my pre-medical school studies, ‘research’ was synonymous with experiments, always practical and sometimes quantitative, and this perspective has continued to be reinforced and dominate throughout my surgical career. This predilection for quantitative research undoubtably affected the lens through which I viewed and reviewed the existing feedback literature, and I the reason the first question I ask is ‘what’ is the effect of feedback models on the efficiency and quality of psychomotor task performance. Furthermore, as a motivated and inquisitive teacher, I have hopes of furthering the body of knowledge, particularly in relation to the coalescing model of dialogic feedback and this motivation should be recognised when considering my involvement in data collection, in both quantitative and qualitative arms of the study.

In relation to the medical students that might be recruited to the study, I am an Orthopaedic Surgery Registrar and Clinical Teacher and may have been encountered in either of those roles. I would suggest that any student with a negative impression of me, in relation to either of those two roles, is unlikely to volunteer for a research project run by me. Therefore, it seems likely that participants are likely to have a neutral to positive bias towards me as a surgeon or teacher, which may affect their engagement in the study, motivation and even response to feedback.

In the latter stages of research, my perspective as a surgeon and MD candidate should be attended to when reviewing the reported outcomes of the study. Whilst robust and objective design of the quantitative section of the study should protect the integrity of the results to a greater degree, the nature of qualitative research is by nature inductive and the process cannot occur without the researcher, and the researcher cannot come without their own agency and context.

However, 'preconceptions are not the same as bias, unless the researcher fails to mention them' (Malterud, 2001, p. 484). Indeed, Attia and Edge (2017) encourage us to move away from the battle with removal of bias, to embracing research within which we are an integral part and are in a position to understand and interpret the complex relationships at play. If this view is supported, who better to conduct a research study investigating the relationship between feedback and psychomotor skill that a highly-motivated surgeon and medical teacher?

Therefore, I propose that the 'solution' to this position is one of acknowledgement, challenge and balance. My perspectives and beliefs have been shaped by previous experiences and I acknowledge that they will interact with all stages of this research process. However, decisions and interpretations should be challenged and balanced by both my own and external perspectives, such that they are not unduly shaped by a single, insider researcher's thoughts and attitudes. In this process, I believe reflective discussion with peers and critical discussion with supervisors will play an important role.

4.5 THE PSYCHOMOTOR TASK

In order to be suitable for a study investigating the effect of feedback on psychomotor task performance, the task used in this study, should meet certain criteria. It should have measurable outcomes of performance, be reproducible, and have controlled participant access. The task must be teachable, authentic to surgery, and allow for improvement in participant performance. There were also pragmatic considerations and it must be feasible in relation to available resources and funding.

For quantified comparison of performance, between different task repetitions and participants, a task with quantifiable measurements of performance was needed. The previously completed literature review (Chapter 2) illustrated the use of quantitative measures of efficiency and quality in relation to task performance within experimental feedback studies. Including measures of both efficiency and quality in the study and task design was desirable, as it would allow for a more detailed and holistic comparison of performances. Another key consideration was whether to employ objective or subjective measures of performance. The previous review illustrated that whilst there is evidence within the literature to suggest that feedback is associated with improved performance (section 2.3.5), analysis of this evidence also suggests that it is more difficult to prove this association via objective measures of efficiency than it is via subjective measures of quality. In order to more stringently explore the quantifiable effects of the feedback model on task performance, and to reduce any effect of investigator bias, I concluded that the use of objective measures of performance was valuable.

Reproducibility of task conditions is key to ensuring a valid trial. The starting conditions, intra-task challenges and end-point must be consistent between repetitions. Replicating these conditions as accurately as possible will increase the sensitivity of measured performance differences and the validity of conclusions drawn from them. Reproducibility is increased by a high degree of control of the task variables and by having a consistent end-point that must be achieved. However, ideally this would be balanced with some tolerance of variation in relation to how the task is approached by participants. To have a task which must be performed in exactly the same fashion each time, repeating in sequence precisely the same steps, will limit variation in participant performance, remove an element of decision making and executive thinking on their part and blunt the sensitivity of inter-performance differences. At its extreme, the task would be reduced to a binary conclusion; performed correctly or not.

The range of potential psychomotor tasks that could have be used in this study is almost endless. However, if the participants were exposed to similar tasks or were able to access the task freely out with the study, this would confound the study results. Therefore, it is vital that access to the task is easily and reliably controlled for each participant for the duration of their study participation.

The study is based within medical education and, therefore, using a task genuine to a medical or surgical setting would increase the authenticity of the study for medical participants and potentially their commitment to the study.

The psychomotor task featured in this study had to be something about which I was able to construct meaningful feedback. The task had to involve the use of skills I possess and was able to teach. It needed to be a task in which I was more experienced than the participants, otherwise my position as instructor and author of feedback might be undermined or even redundant. In the selection of a task, my position as a surgical trainee was relevant, as well as consideration of the experience of potential participants.

Complexity of the task was an important factor in that is directly relevant to how achievable the task would be for participants and how sensitive it would be in illustrating variation in performance. Complexity of task had to be considered within the context of pre-study participant competence, the length of the study and number of task repetitions completed. Whilst some variation in baseline exposure, competence and natural ability of the participant would be inevitable, huge variation within the participant population would have led to difficulties and inaccuracies in comparison. If the task was too simple, it could be mastered too easily and a potential ceiling effect with a plateau in quantifiable performance measures could be encountered. However, if the task was too complex, participant proficiency may improve too slowly for change to be quantified. The desired level of complexity was specific to the type of task chosen and accurate judgement required pilot investigations.

As a postgraduate research student at the Centre for Medical Education, University of Dundee, I had valuable access to Dundee Medical School facilities on the Ninewells Hospital site. This included resources at the Clinical Skills Centre, Cuschieri Skills Centre and Tayside Orthopaedic and Rehabilitation and Technology (TORT) Centre. Although I may have been able to apply for additional funding to cover additional costs related to the experimental study design, it was pragmatic to first explore the psychomotor tasks accessible at these existing resources before looking further afield.

4.5.1 Potential psychomotor tasks

4.5.1.1 Bench model suturing and knot tying

Wound suturing and knot tying are basic surgical skills that have been used previously in the investigation of the effect of feedback (Farquharson, 2013; Roger, 2000; Boehler, 2006; Xeroulis, 2007; Porte, 2007; Rafiq, 2008) and I considered a suturing-based bench model task for this study. Outcome measures could include the time taken to close a wound of a set length, quantifying instrument travel (economy of movement) via an optical tracker and quality of the knots tied. Task reproducibility would be achieved by using identical suturing models and, furthermore, suturing is something I have significant experience in and is authentic to the medical setting.

However, on deeper consideration of a suturing task, several additional issues were raised. Objective measures of efficiency are possible but objective measures of accuracy are practically more difficult, which is borne out in the reporting of measures of only efficiency in several relevant studies (Backstein et al, 2004; Boehler et al, 2006; Farquharson et al, 2013). It would have been possible to control the length of the wound but to objectively measure the adequacy of closure (degree of gape or tension) would be very difficult. In order to control proficiency of closure, one could specify the number of sutures required but this would limit participant decision making and may make performance susceptible to a ceiling effect. For these reasons, a wound suturing and knot tying task was not deemed ideal for use in the study.

4.5.1.2 Removal of a mass

I considered potential amendments of the suturing task to increase its complexity and provide additional measures or performance. The removal of a mass requires an increased number of skills and greater executive thinking than the simple closure of a wound. Instrument travel during skin incision and dissection as well as total time taken would provide measures of efficiency. The mass would be palpated and skin marking of the size of the mass via the palpated border, plus measurement of the total volume removed, mass plus normal tissue, would provide quantifiable measures of accuracy. Skin closure could form part of this task, without having to be the sole aim. In principle, this task certainly meets the criteria of being a teachable, authentic skill and it would be possible to control participant access.

On the basis that a reusable, commercially produced model was not available, I consulted with the technicians at the Cuschieri Skills Centre for advice in how I might construct a suitable bench model. They have considerable experience in producing models for surgical skills courses and producing the models myself would reduce costs and control task conditions. I trialled production of a layered model which consisted of neoprene for synthetic skin and a mass lying within deeper adipose tissue. The models were constructed using layers of foam of differing densities for the adipose tissue and the mass was constructed of a gelatinous substance encased in latex. Different substances were trialled for the mass, such as modelling putty and PVA glue. The mass was buried within the adipose tissue and adhered to its surrounding tissue to increase the difficulty of removal.

Trials of these models revealed several issues with their practical use. They were time consuming to produce and it was difficult to produce exact replicas. It was difficult to quantifiably assess important factors such as care of instrument use and quality of dissection using summative measures such as weight of removed tissue. The overwhelming 'homemade' feel of the model threatened face validity of the task. For these reasons, this task was not deemed feasible or desirable for use in the study.

4.5.1.3 Simulated laparoscopic tasks

Completion of laparoscopic tasks, an example of minimal access surgery, requires exhibition of several generic surgical skills, including bimanual instrument manipulation and dexterity, depth perception, three-dimensional space appreciation plus planning and decision-making abilities. As such it shares significant similarities to arthroscopy, minimal access joint surgery undertaken in orthopaedic practice. However, unlike arthroscopic surgery, simulated laparoscopic surgery training models are commonplace and easily accessible. Laparoscopic tasks have been previously used in the investigation of feedback (Judkins et al, 2006; O'Connor et al, 2008; Boyle et al, 2011; Strandbygaard et al, 2013; Paschold et al, 2014; van Sickle et al, 2007; Kannappan et al, 2012).

The Cuschieri Skills Centre has two different simulation models: box model trainers and virtual reality trainers. The former utilises a static camera placed within an abdominal cavity model, which is connected to a computer monitor, and standard laparoscopic instrumentation introduced into the abdominal cavity via portals. Box trainers are highly versatile as a wide range of tasks – moving pegs on a board, stacking cubes, cutting shapes – can be incorporated into the exercise. Virtual reality models consist of a monitor, instruments and specialised software installed on a PC. The tasks available for use with the virtual reality models are restricted and dependent on the specific model used but the great advantage is that task conditions can be perfectly controlled and certain performance data is collected automatically.

The LapSim® (Gothenburg, Sweden) model was available for use in the Cuschieri Skills Centre. This PC-based virtual reality system consists of a 21-inch monitor and a laparoscopic 'Basic Skills' interface module with two instruments and a foot-switch. Ten different tasks were available: Camera navigation, instrument navigation, coordination, grasping, cutting, clip applying, suturing, handling intestines and fine dissection. The difficulty of these tasks can be varied, from settings which make the task most simple (level 1) to those that make it most complex (level 3). For the completion of certain tasks, different instruments are required and the instrument can be changed virtually including a laparoscopic grasper, clip applier, scissors, diathermy scissors and suction. A pedal, which is operated by the subject, doubles as a diathermy or suction device depending on which instrument has been selected.

There were key advantages of using a laparoscopic simulator task. The LapSim® simulator automatically collects objective, quantitative performance data relating to both the efficiency and quality of performance. The chosen task can be exactly replicated on multiple occasions aiding reproducibility. Access to the Cuschieri Skills Centre is restricted and, therefore, controlling participant access during the course of the study would be possible. The skills required for these tasks are within my area of expertise and practice as a surgical trainee, giving me the authority and experience required to teach these skills. Whilst use of virtual reality may cause concern regarding authenticity, use of these simulators is widespread within medical surgical training and, therefore, I suspected they would be acceptable to participants. The variety of tasks available, combined with adjustable difficulty levels, gave me confidence that the correct complexity level could be selected to allow task completion whilst avoiding a ceiling effect in performance. Furthermore, there were four LapSim® models available within the Cuschieri Skills Centre and the centre was happy for me to use a model for study purposes.

For these reasons, I first explored and then later selected a virtual reality laparoscopic simulator task for use in this study.

4.6 THE FEEDBACK INTERVENTION

This study is specifically concerned with the effect of the *form* and *nature* of the feedback, not the content. Therefore, it is important for the validity of the study that the content of the feedback should be similar between groups, such that this could not be a confounding factor.

Intended characteristics of both information transfer and dialogic feedback is described in Section 4.2 (Box 4). These would be used as reference for the creation of feedback forms, which would be used to structure the feedback sessions featured in the study. The intention is that these forms could act as an aide-mémoire, promoting adherence to the feedback type, and standardising the content and quality of the feedback given.

4.6.1 Information transfer feedback

A common source of feedback in medical training is Work-place Based Assessments (WBAs). Although primarily concerned with assessment, their secondary purpose is to provide structured feedback to the learner. A Procedural Based Assessment (PBA) is a type of WBA concerned with the appraisal of performance of a specific surgical procedure. PBAs were developed by the Joint Committee on Surgical Training and are one of the WBAs that form the battery of evaluation tools featured in the Intercollegiate Surgical Curriculum Programme. Separate PBAs exist for each different surgical procedure of interest during UK-based surgical training. PBAs contain five subsections: *Pre-operative planning, pre-operative preparation, exposure and closure, inter-operative technique* and *post-operative management*. Each subsection is further broken down into specific elements and ratings are applied by the tutor to each of these elements: Not seen during this episode (N), development required (D), or satisfactory (S). The 'Intra-operative technique' section relates to feedback and assessment specifically concerning the performance of the procedure and contains *'Global' (G)* and *'Task specific' (T)* elements. Each PBA incorporates a *'Global summary'*, in which the performance of the procedure is graded from level 0 (least proficient) to 4 (most proficient). The feedback given during PBA completion is inherently information transfer-type feedback. It is tutor driven and dependent and the learner role is passive. The content of the feedback focuses on the outcome of actions. The PBA forms promotes highly structured and detailed feedback.

4.6.2 Dialogic feedback

The purpose of the dialogic feedback sessions is to encourage tutor-facilitated participant self-analysis of performance and to encourage participant self-regulation with adjuvant tutor observations and direction. Therefore, the purpose of the dialogic feedback form is to provide a shortcut to the notetaking of detailed and accurate researcher/tutor observations. These observations support well informed discussion during feedback sessions, with opportunities to adapt and improve performances recognised and successes reinforced. As both the information transfer and dialogic feedback forms allow opportunity for detailed and specific researcher/tutor note taking, the content and quality of the feedback are standardised between the information transfer and dialogic feedback groups, whilst the format for the feedback itself differs. The dialogic feedback form is completed by the tutor\(researcher) as deemed helpful for notetaking during the task performance and used as an aide-mémoire for content during explorative discussion during feedback sessions. The dialogic feedback sessions will be less prescriptive than the information transmission sessions; the form itself will not be directly referred to or ‘worked through’ in the same way. The focus of the dialogic feedback form is to allow discussion of process rather than outcome and judgement.

CHAPTER 5: PILOT STUDY ONE

5.1 PURPOSE OF THE PILOT STUDIES

Two pilot studies were conducted as part of the planning process for the main study. These helped inform the final study design and they gave me valuable practical experience in running an experimental study.

The particular focus of pilot study one was the practicalities of running an experimental cohort study. Specifically, I wanted to explore the feasibility of use of a simulated laparoscopic task and to gain some experience in the practicalities of running an experimental study, and the use of participant information sheets, consent forms, and data collection.

The focus of pilot study two was the intervention, that is, the feedback itself. Specifically, I wanted to explore how the different types of feedback can be orchestrated and articulated and to gain a better understanding of how the two, different feedback models would be expressed and created. Additionally, I wanted to build on the practical experience gained in the first pilot study in relation to running an experimental study to ensure the most robust protocol and materials possible for the final study design.

5.1.1 Aims of the study

Two pilot studies were conducted as part of the planning process for the main study. These helped inform the final study design and they gave me valuable practical experience in running an experimental study.

The particular focus of pilot study one was the psychomotor task. Specifically, I wanted to explore the feasibility of use of a simulated laparoscopic task and to gain some experience in the practicalities of running an experimental study, including the use of participant information sheets, consent forms, and data collection.

I synthesised specific questions to ensure that the first pilot study design and analysis would yield relevant information:

- Which laparoscopic tasks specifically would be suitable?
- What task settings give the appropriate level of difficulty to participants? What is the approximate spread of measured indicators of performance?
- Will it be necessary to have different levels of difficulty to show progression or deterioration in performance and differentiate participant performance?
- What is the effect of repetition without feedback on performance?
- Is the setting of the study appropriate?
- Are the forms used to collect the data appropriate?

5.1.2 Ethical approval

Ethical approval for this study was sought prior to commencement of the pilot studies. The University of Dundee Research Ethics Committee (Application 14134) granted permission on 29th October 2014. (Appendix A)

5.2 METHODS

5.2.1 Study design

5.2.1.1 Participants, recruitment and randomisation

Second year undergraduate medical students at the University of Dundee were recruited to this pilot study via an invitational email to their year group. In return for participating in the pilot study, the students were offered a certificate of participation for their undergraduate portfolio.

Nineteen second year undergraduate University of Dundee medical students volunteered for the study.

As the main purpose of this study was to gain information about the suitability of the task chosen and the practicalities of running an experimental study, I was not concerned specifically about controlling group size nor the randomisation process. Participants were allocated into three groups based upon the day that they were able to attend:

- Day one: *Control group*
- Day two: *Information transmission feedback*
- Day three: *Dialogic feedback*

5.2.1.2 Study setting

The study was based at the Cuschieri Surgical Skills Centre, Ninewells Hospital. The LapSim® simulators are based in one of the rooms within the centre and, as the medical students are based in the Medical School also within Ninewells Hospital, access to this venue would be convenient.

5.2.1.3 Study protocol

Figure 16 refers to the study protocol employed in the first pilot study. The terms 'information transfer' and 'dialogic' feedback will be explained later in the chapter. Each study visit was structured as per the 'session components' described.

Figure 16: Pilot study one protocol

Session component	**Participant group**		
	Control group, n = 10	*Information transfer feedback group, n = 5*	*Dialogic feedback group, n = 4*
Orientation	Study information sheet issued; Consent form issued / signed; Pre-study data collection sheet completed		
Explanation of task	Written task instructions given, tutor question answering, tutor demonstration.		
Task performance	**Task 1 performance 1 (P1)**		
Task performance	**Task 1 performance 2 (P2)**		
Feedback session 1	-	Structured information transfer feedback episode	Dialogic feedback episode
Task performance	**Task 1 performance 3 (P3)**		
Feedback session 2	-	Structured information transfer feedback episode	Dialogic feedback episode
Task performance	**Task 1 performance 4 (P4)**		
Debrief	Participants thanked and any questions answered. Participants depart.		

Each participant completed a single 30-minute study visit. The task was performed a total of four times by each participant. Following task performance two and three, the feedback groups received or took part in an episode of feedback. No feedback took place between performances one and two to prevent confusion between the effect of feedback and initial acclimatisation of the naïve learner to the instruments and task. The control group (n=10) were included to investigate the effect of repetition on task performance without any external feedback. This group completed the task four sequential times without participation in feedback sessions.

5.2.1.4 Orientation

At the start of the study visit, participants were provided with a study information sheet (Appendix B) and informed consent (Appendix C) was obtained. Pre-study participant data was also collected (Appendix D). This included age, sex, dominant hand, self-rated confidence level, self-rated experience level, description of experience and participant confidence in tutor (researcher).

5.2.1.5 Participant task instructions

Participants received standardised task instructions via a printed instruction sheet (Appendix E) and were asked to read this instruction sheet at the start of the study visit and there was an opportunity for clarification of the instructions prior to the first task performance. The researcher then answered any participant questions prior to a brief demonstration, which illustrated the use of the four instruments, grasping the vessel, vessel rupture due to excessive tension, applying a clip, the use of suctions and use of the scissors.

5.2.2 The psychomotor task

As previously mentioned, the LapSim® has ten potential tasks that could be used. The tasks vary in terms of their complexity, such as the degree of tactical thinking required, number of task subcomponents, accuracy of manipulation required, and number of different instruments required during task.

The 'clip applying' task was chosen for the first pilot as it appeared to be the most sophisticated task and, therefore, offered a challenge to the participants in relation to both dexterity and in metacognitive terms.

5.2.2.1 Clip applying task technique

The aim of the clip applying task is to ligate a vessel lying in the abdominal cavity before then dividing the vessel (Figure 17). The vessel is ligated by placing a correctly applied clip in each of the green target areas (one on the left and one on the right side of the vessel) and the vessel is then divided in the blue cutting target area (in the centre of the vessel) using the scissors.

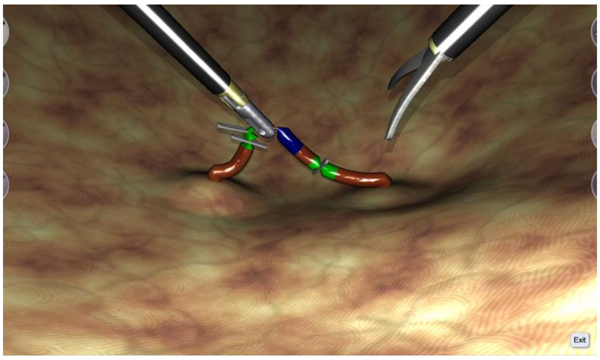


Figure 17: The vessel in the abdominal cavity with green clip target areas and blue cutting target

The blood vessel is susceptible to stretch damage. As the blood vessel is stretched, it turns increasingly red (Figure 18). If stretched past its tolerance level, it will rupture and bleed. If this occurs, the clip target areas will disappear and are no longer be relevant to the task completion. The number of missed or incomplete target areas are counted at this stage. If the vessel starts to bleed, the task can only be completed by placing a clip on each side of the vessel, below the rupture to stop the bleeding (Figure 19). In order for the LapSim® to be satisfied that the clip has been correctly applied, both jaws of the clip must be visible during application. This encourages a safe clip application technique. Correct clips can be identified as they show an equal length of clip on either side of the vessel (Figure 20). They can be differentiated from 'badly applied clips', which appear asymmetrical (Figure 21). Any dropped clips (Figure 22) plus any volumes of blood greater than 100ml (Figure 19) must be evacuated using the suction device before the task is deemed completed. Although it does not affect task completion, the screen will turn red to alert participants that undesirable damage to the abdominal wall is being sustained via instrument intrusion (Figure 23). The screen will return to normal once the pressure is relieved. The programme will end the task automatically once all criteria for completion are satisfied.

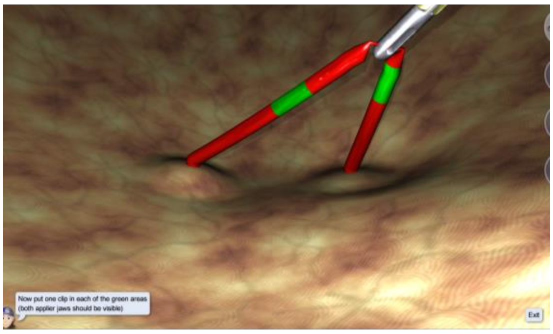


Figure 18: The vessel is stretched and has become increasingly red

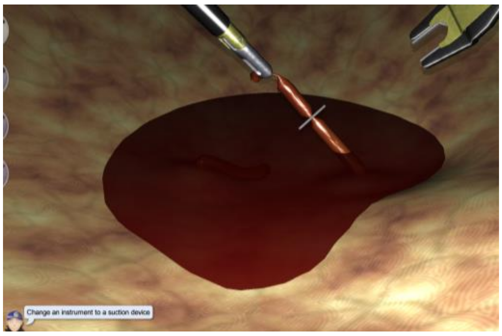


Figure 19: The vessel has ruptured and bled. The left side of the vessel has been ligated with a clip.

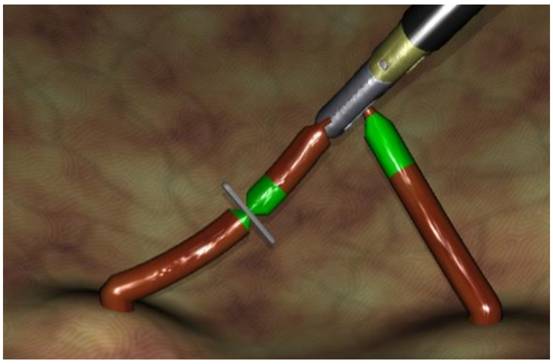


Figure 20: A correctly applied clip

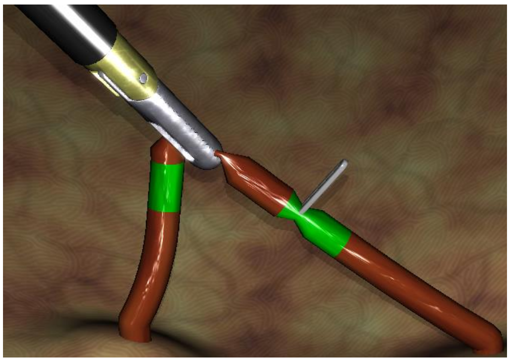


Figure 21: An incorrectly applied clip

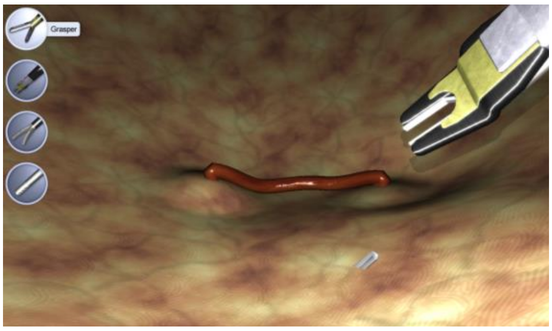


Figure 22: A dropped clip on the abdominal floor

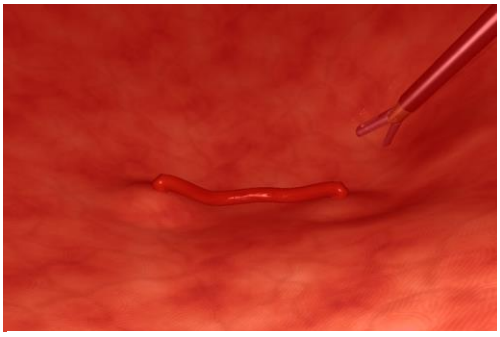


Figure 23: The red screen signifies abdominal floor damage via instrument intrusion

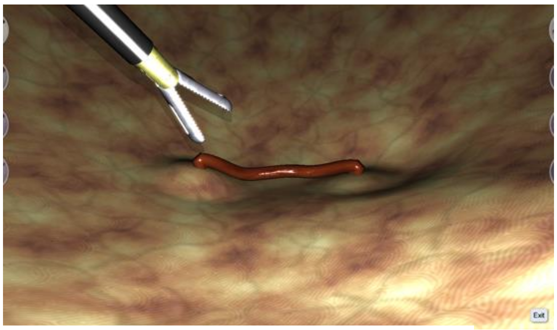


Figure 24: The grasper

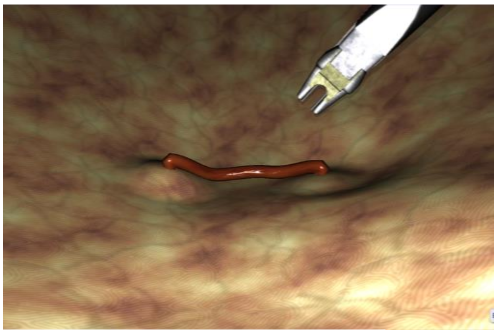


Figure 25: The clip applier

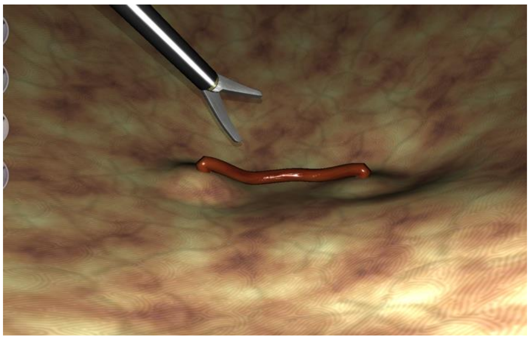


Figure 26: The scissors

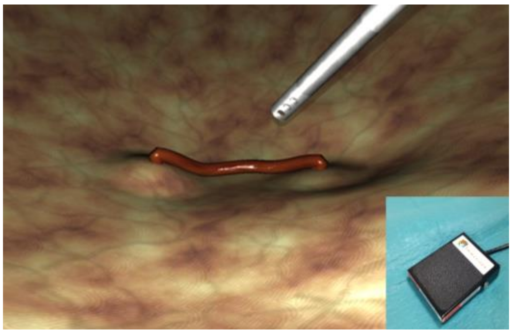


Figure 27: The suction

Figure 28: Withdrawal of hand instruments

Figure 29: Handles of hand instruments in open (left) and closed (right) position for changing of instruments and closure of grasper, clip applier and scissors

Figure 30: Cogwheel for rotation of instrument tips

Use of at least three (grasper, clip applier and scissors, Figures 25-27 respectively) and potentially a fourth (suction, Figure 28) instrument is required to complete the task. The foot pedal (Figure 27) is pressed to activate the suction device. These instruments are virtually changed by the participant by withdrawing the hand instrument (Figure 28), closing the handles (Figure 29) repeatedly until the desired instrument is selected, before reinserting the instrument. You can rotate the end of the instrument by using the cogwheel on the handles (Figure 30).

The exact method employed to complete the task can be varied. For example, the left or the right clip can be applied first; the grasper may or may not need to be repositioned after the second clip depending on the position of the applied gasper; the order of clip application and the hand using the scissors can be different with the task still successfully completed.

5.2.2.2 Task setting options

The LapSim® software allows configuration of the task, which allows reproducibility and manipulation of difficulty level. Configurations fall into three broad categories:

Camera options:	At 0° the camera faces the vessel straight on. By entering either negative or positive values, the camera can be positioned to look at the vessel either from the left or right.
Vessel options:	The size of both clipping and cutting target areas (in millimetres) can be set, as can the stretch sensitivity of the vessel (low, medium or high), which dictates how easily it is ruptured.
Bleeding options:	The flow rate of blood from a torn vessel can be set (in litres/minute). The programme includes a ‘spontaneous bleeding’ option, in which the vessel will spontaneously bleed regardless of stretch.

5.2.2.3 Selected task settings

Box 5 shows the settings selected for the purposes of this pilot study. These settings were selected as a starting point based on both the programme 'automatic' settings and informal investigator trials.

Box 5: Pilot one LapSim® 'Clip application' task settings	
Camera angle	-20 degrees (vessel slightly oblique to viewer, left side closer)
Clip target area size (mm)	4
Cutting target area size (mm)	4
Stretch sensitivity	Low
Spontaneous bleeding	Off
Bleeding flow rate (L/min)	1

5.2.2.4 Task measures of performance

The LapSim® collects eight separate measures of performance in relation to the clip applying task as detailed in Box 6. These measures can be used to assess participant efficiency and accuracy.

Box 6: Pilot one LapSim® 'Clip application' measures of performance	
Total time taken	*(TTC, in seconds)*
Right instrument path length	*(RIPL, in centimetres)*
Left instrument path length	*(LIPL, in centimetres)*
Number of incomplete (missed) target areas	*(ITA, raw number)*
Number of badly placed clips	*(BPC, raw number)*
Number of dropped clips	*(DC, raw number)*
Maximum blood vessel stretch	*(BVS, expressed as percentage of maximum tolerated)*
Blood loss	*(BL, in millilitres)*

5.2.3 The feedback intervention

5.2.3.1 Information transfer feedback

The information transfer feedback form (Figure 31, below, and Appendix F) was adapted from the 'Intra-operative technique' section of the 'Generic Laparoscopic Hernia Repair' PBA (Appendix G). The global elements address the over-arching skills required for competent task completion: following a logical sequence of actions, careful tissue handling, prompt control of blood loss, appropriate use of instruments, economy of movement and calmly responding to complications encountered. The task specific elements were tailored to the laparoscopic vessel ligation task: grasps vessel carefully, applies clips correctly, cuts vessel safely and retrieves dropped clips. Space for free text comments relating to each of the elements will allow the researcher/tutor opportunity to record observations to aid individualised and detailed feedback. The feedback form also incorporates the 'Global summary' section featured in the PBA and space for recording the length of the feedback session.

Date:

Information transfer: Pilot 1
Appendix E

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study: Information transfer group
Feedback checklist

Participant number: Task number: Time taken for FB:

	Component	**Rating**			**Comments**
G1	Follows and agreed, logical sequence or protocol for the procedure	N	D	S	
G2	Consistently handles tissue well with minimal damage	N	D	S	
G3	Controls bleeding promptly	N	D	S	
G4	Uses instruments appropriately and safely	N	D	S	
G5	Proceeds at an appropriate pace with economy of movement	N	D	S	
G6	Deals calmly and effectively with unexpected events and complications	N	D	S	
T1	Grasps vessel carefully with grasper to allow clip application	N	D	S	
T2	Applies clips correctly to either side of vessel	N	D	S	
T3	Cuts vessel safely with scissors	N	D	S	
T4	Retrieves dropped clips	N	D	S	

Level of performed task		
0	Unable to complete task	
1	Able to complete task with significant difficulty in several components	
2	Able to complete task with mild-moderate difficulty in several components	
3	Able to perform all elements of task with mild difficulty in only some components	
4	Able to perform all elements of task, exhibiting high skill level	

Figure 31: Information transfer feedback form

This feedback form will be completed in full during each information transfer feedback session. The tutor (researcher) will work through each of the global and task specific components, providing a participant with a rating for each component and discussing the comments relevant to each task component. The feedback session will conclude with the tutor providing an overall rating of the performance via the global rating. These feedback sessions will be highly structured and the tutor will lead discussion.

5.2.3.2 Dialogic feedback

The dialogic feedback form (Figure 32, below, Appendix G) incorporated five of the global and task specific items of the information transfer feedback form to promote content similarity. ‘Prompt control of blood loss’ was removed as it is itself a feature of ‘response to complications’. The language used for the elements was altered thus that the emphasis was on observation of behaviour (‘tissue handling’) rather than assessment compared to a desired performance level (‘consistently handles tissue well with minimal damage’). An element relating to evidence of self-regulation was added so that behaviour specifically relating to this could be more easily noted and then highlighted during feedback. The N, D and S ratings were removed and replaced with two columns for researcher/tutor observations; those relating to systematic or tactical awareness and those relating to technical skills. The ‘global summary’ was removed as it is incongruent with learner driven feedback.

Date:

Dialogic: Pilot 1
Appendix G

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study: Dialogic group
Feedback model

Participant number: Task number: Time for FB:

	Component	**Systematic / tactical awareness**	**Technical skills**
G1	Logical sequencing		
G2	Tissue handling		
G3	Use of instruments		
G4	Pace with economy of movement		
G5	Deals calmly and effectively with unexpected events and complications		
G6	Evidence of self-regulation		
T1	Use of grasper to allow clip application		
T2	Applies clips correctly to either side of vessel		
T3	Control of bleeding		
T4	Cuts vessel safely with scissors		
T5	Minimises dropped clips / method retrieval		

Figure 32: Dialogic feedback form

The purpose of the dialogic feedback form was an aid to accurate but efficient collection of tutor (researcher) observations relating to participant performance, such that it could provide useful and detailed content during feedback sessions, but did not unduly detract from keen observation. It is not intended that these forms be discussed in full or in a stepwise fashion, as the case with the information transfer model; they should be regarded purely as tutor notes. Instead, each feedback session will start with the tutor (researcher) inviting reflection from the participant, before the discussion of co-created of dialogue in relation to task performance. It is not intended that these sessions will be rigidly structured, but instead content will be largely drawn from the participant, with clarification and direction from the tutor (researcher) as required.

5.3 RESULTS

The raw data and tabulated results and analysis for pilot study one are available for viewing in the accompanying ‘Data disk’ (filename ‘Pilot study 1 raw data’ and ‘Pilot study 1 results’ respectively). Brief statistical analysis was performed with respect to each task performance 1 – 4 (P1 – P4, see Section 5.2.1.3 ‘Study protocol’) and the quantitative performance measures: time taken to complete the task (TTC), volume of blood loss (BL), number of incomplete target areas (ITA) and number of badly placed clips (BPC). These were regarded as markers of efficiency (TTC and BL) and markers of accuracy (ITA and BPC).

The chief purpose of this analysis was to explore the validity of the task chosen and to ascertain the success of the difficulty level chosen. The mean, the minimum and maximum values were calculated for each performance measure, for each group, in each performance. For the task and difficulty level to be deemed suitable for the study, the performance measures achieved should allow for differentiation in participant performance, without a floor or ceiling effect observed.

5.3.1 Time taken to complete (TTC)

The mean time taken to complete the task (TTC) for each group at each performance repetition, P1 – P4, is shown below in Table 12. Shown in Table 13 is the minimum and maximum TCC for each group in each performance, which illustrates range and variation in performance.

Table 12: Average time taken to complete task (TTC)

	Mean time taken to complete task (TTC) (seconds)			
Group	**P1**	**P2**	**P3**	**P4**
Control (repetition only)	321	205	200	165
Information transmission feedback	381	279	264	304
Dialogic feedback	346	219	212	232

Table 13: Minimum and maximum time taken to complete task (TTC)

Group	Range in time taken to complete task (TTC) (seconds)							
	P1		**P2**		**P3**		**P4**	
	Min	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*
Control	172	602	155	476	85	433	166	442
Information transmission	165	548	132	455	53	457	70	300
Dialogic	277	479	115	373	99	418	55	535

5.3.2 Volume of blood loss (BL)

The mean blood loss sustained by each group at each performance repetition, P1 – P4, is shown below in Table 14. Shown in Table 15 is the minimum and maximum blood loss for each group at each performance, which illustrates range the range of this performance measure.

Table 14: Average blood loss (BL) by group and task performance

Group	Mean volume of blood loss (BL) in litres			
	P1	**P2**	**P3**	**P4**
Control (repetition only)	1.0	0.4	0.3	0.2
Information transmission feedback	1.9	1.3	1.3	0.8
Dialogic feedback	1.8	0.7	0.5	0.6

Table 15: Minimum and maximum blood loss by group and task performance

Group	Range in blood loss (BL) in litres							
	P1		**P2**		**P3**		**P4**	
	Min	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*
Control	0.00	1.99	0.00	1.14	0.00	0.63	0.00	0.49
Information transmission	0.52	6.25	0.2	3.22	0.21	2.93	0.00	1.33
Dialogic	1.04	3.13	0.22	1.30	0.19	0.96	0.10	1.39

5.3.3 Number of incomplete target areas (ITA)

The mean number of incomplete target areas (ITA) for each group at each task performance repetition, P1 – P4, is shown in Table 16 below. Shown in Table 17 is the minimum and maximum ITAs for each group at each performance, which illustrates the range of this performance measure).

Table 16: Average incomplete target areas (ITA)

	Mean number of incomplete target areas (ITA) (Range 0 – 3)			
Group	**P1**	**P2**	**P3**	**P4**
Control (repetition only)	1.7	1.7	1.5	1.4
Information transmission feedback	1.8	1.8	2.0	1.6
Dialogic feedback	2.0	2.0	2.0	1.8

Table 17: Minimum and maximum incomplete target areas task by group and task performance

	Range in blood loss (BL) in litres							
Group	**P1**		**P2**		**P3**		**P4**	
	Min	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*
Control	0	3	0	3	0	2	0	2
Information transmission	1	2	1	2	2	2	0	2
Dialogic	2	2	2	2	2	2	1	2

5.3.4 Number of badly placed clips (BPC)

The mean number of badly placed clips (BPC) placed by each group at each performance repetition, P1 – P4, is shown in the Table 18 below. Shown in Table 19 is the minimum and maximum number of badly placed clips for each group at each performance, which illustrates the range and variation in this performance measure.

Table 18: Average number of badly placed clips (BPC)

	Mean number of badly placed clips (BPC)			
Group	**P1**	**P2**	**P3**	**P4**
Control (repetition only)	1.4	3.5	4.5	2.6
Information transmission feedback	1.0	1.2	2.8	3.0
Dialogic feedback	1.25	1.0	0.0	0.0

Table 19: Minimum and maximum number of badly placed clips by group and task performance

	Range in number of badly placed clips (BPC)							
Group	**P1**		**P2**		**P3**		**P4**	
	Min	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*
Control	0	3	0	17	0	15	0	7
Information transmission	0	4	0	3	0	5	0	5
Dialogic	0	3	0	3	0	0	0	0

5.4 DISCUSSION

5.4.1 Summary

- The setting of the study was appropriate but issues with access to the room housing the LapSim® models were identified as a potential problem with longer studies.
- Use of the laparoscopic simulator allowed task allowed accurate data collection and objective quantitative data suggested variation between participants.
- More detailed task instructions would be required in future studies due to the variation in participant understanding encountered in this pilot.
- The clip application task could be successfully performed by participants and as no plateau in performance was encountered, the task settings were acceptable and multiple difficulty levels do not appear to be required for further pilot tests.
- Use of tailored feedback forms helped orchestrate the different models of feedback and were a useful aid in the accurate recording of observed behaviour, which provided detailed content for feedback sessions.

- On average, participants in all three study groups performed the task more quickly in the final task performance than in the first task performance but did not reach, or even approach, a floor effect.

5.4.2 Study design

5.4.2.1 Participants, recruitment and randomisation

Second year medical students were readily recruited to the study and email would appear to be a valid method of recruitment. Randomisation by date of first visit was acceptable for this initial pilot study but as participant availability might relate to timetables and student groups, an alternative method of randomisation was sought for the second pilot study.

However, five additional students volunteered for the study but later became unable to make their appointment time and thus did not take part. This is a relatively high non-starter rate and such a rate may pose a significant difficulty to sufficient recruitment to the final, larger study. Reasons for non-starting may be attributed to participant or organisational factors. Participant factors might have included deciding against, forgetting or becoming unable to attend with little to no notice to reschedule. Information from these potential participants suggested that the latter was the most common issue, and this relates to the organisational issue identified. Organisationally, there was an issue in lack of flexibility of available appointment times, both in relation to day (the study ran over only three days) and time (there were certain times during the day that the room was not available). Whilst the duration of data collection for this pilot study was a conscious design choice made for practical reasons, the limitations in times available during those days was due to other activities occurring in the room through which the LapSim® machines were accessed in the Cuschieri Centre. This prompted me to explore an alternative location for the second pilot study.

5.4.2.2 Orientation

The study information sheet, consent form and pre-study data collection sheet all appeared to perform well in this small pilot study. Participants seemed to understand and be satisfied with what was being asked of them during the study visit.

5.4.2.3 Task instructions

The success of using written instructions with opportunity for questions to clarify and a tutor demonstration was limited. Overall, the method was inefficient and variation between participants affected the quality of the pilot study with potential implications for study validity.

The time participants spent reading the instruction sheet varied and I suspect not all participants read them fully. The volume and detail of participant questions varied, from many detailed questions to none whatsoever. Although the demonstration followed the same basic pattern each time, the length varied slightly due to participant questions. I was concerned that variation in participant understanding prior to starting may impact the validity of results.

5.4.3 The psychomotor task

5.4.3.1 Task selection and settings

The use of the simulated laparoscopic vessel ligation task appeared to work well in this study. All participants were able to complete the task and the automatically collected quantitative measures of efficiency and accuracy illustrated differences between intra-and inter-participant performances.

The researcher was consistently able to complete the task, fulfilling all target areas and using the minimum number of clips, in 20-30 seconds. The results of this study show that this level of performance was not achieved by the participants in this single-visit study and further improvement in performance could be achieved and measured. The results support the use of the difficulty settings chosen and suggest that use of different levels of difficulty may not be needed to show progression in skill and performance.

5.4.3.2 Quantitative measures

The quantitative measures of task performance selected for inclusion in this pilot study were accurately collected and there was no missing data. The range of results in relation to time taken to complete the task (TTC) and blood loss (BL) was high and it was identified that this may cause some issues in relation to the use of this performance marker for statistical design purposes and inter-group comparisons. However, this had to be investigated further in a pilot study with more rigorous testing of the feedback intervention before final statistical plans could be made.

5.4.4 The feedback intervention

Whilst the intervention was not the primary focus of this pilot study, it afforded an opportunity to test the different feedback models and the use of the forms constructed.

In brief (researcher) reflection, it was possible to instigate and orchestrate the different types of feedback. In the information transfer group, feedback was highly structured, comprehensive, less comprehensive and tutor (researcher) driven. Learner (participant) input was limited, in both range of contribution (largely acknowledgement of receipt of feedback) and volume. Feedback in the dialogic group differed in that the contributions of researcher and participant were more equal. The feedback was less structured and more exploratory and conversational. However, the level of detail in relation to content was similar.

This pilot study highlighted that, as would be expected with a structured task, it was possible to identify specific positive and negative tactical and skill behaviours in the repetitions of task completion across the participant group. This led to the same tutor observation notes being made over and over again. For example, with regards to control of bleeding, a common error was to immediately select the suction device as soon as bleeding started. This would only delay a definitive arrest of bleeding, which required the use of a grasper and clip applying device. A more successful tactic was to avoid suction until visualisation was impaired and instead focus on clip application to the vessel ends. Another error was that if participants did not recognise where the bleeding was coming from (this could be either the end or side wall of a severed vessel) a clip may be placed in the wrong position (too distal) to arrest bleeding. The more successful course of action, if there was any doubt as to the exact site of bleeding, was to place a clip proximally (low down) on the vessel to stem blood flow and arrest bleeding. Furthermore, the vessel commonly ruptured during an attempt to apply the first clip. Therefore, the bleeding would be immediately apparent from the flail side of the vessel but not from the side held in the grasper. A common error would be to drop the held side of the vessel and pursue the other, bleeding, side. The more successful approach is to clip the held vessel, preventing bleeding on that side, before transferring to the other side to arrest bleeding.

Based upon this experience gained from this pilot, I was able to compile these observations and incorporate them into the free-text boxes on each of the information transfer and dialogic feedback forms. The updated versions were used in the second pilot study. The aim was to shortcut the note making required during observing task performance, both for ease of use (maximising opportunity for direct observation) and improving inter-tutor feedback consistency.

5.4.5 Results

5.4.5.1 Time Taken to Complete the task (TTC)

The Time Taken to Complete the task (TTC) shows a general trend of increasing efficiency from performance 1 to 4. The range was relatively high but no floor effect was observed and no performance had to be terminated due to inability of the participant to complete the task. The TTC showed that even in the absence of feedback, repetition of task could and did produce an improvement in efficiency of performance.

5.4.5.2 Blood loss (BL)

Examining the mean blood loss sustained by each group at each performance repetition illustrates an improving trend over the course of the study. Variation, again, was relatively high. No participant managed to consistently achieve zero blood loss, indicating a high rate of vessel rupture. The control group exhibited an improving trend across the duration of the study, again suggesting that even in the absence of feedback, repetition of task could and did produce an improvement in performance.

5.4.5.3 Incomplete target areas (ITA)

The number of incomplete target areas at the end of a performance is a measurement of (in)accuracy of task performance. At performance 4, the mean number of incomplete target areas ranged between 1.4 and 1.8 across the three study groups. As the range of this measure was limited to 0 – 3, the variation in this measure is of course lower. When considering all task performances in all three groups, the modal number of incomplete target areas was 2 (n=19, 75%). Very few participants were able to demonstrate a sustained improvement of performance in relation to this performance measure although there was episodic improvement, which shows that improved performance was possible.

5.4.5.4 Badly placed clips (BPC)

Unlike the other performance measures, range in the number of badly placed clips (a measure of inaccuracy) per task performance was highest in the control group. In the control and information-transfer feedback groups, the mean number of badly placed clips did not follow a decreasing trend, which may indicate that either this is not a valid measure of task performance or that participants did not understand the requirements for placing a ‘good clip’ from reading the task instructions. This performance measure cannot be affected by task settings; the requirements for placing a ‘good’ clip were explained in the task instructions and each clip is judged as either properly placed or not by the LapSim® programme. Interestingly, the dialogic feedback group were able to improve their mean performance in relation to this measure and their third and fourth performances. and there is evidence of a ceiling effect in performances within this group. Regarding all 16 repetitions of the task in this group, the modal number of badly placed clips was 0 (n=12, 75%). This performance measure will be further examined in the second pilot.

CHAPTER 6: PILOT STUDY TWO

6.1 INTRODUCTION

6.1.1 Aims of the study

The purpose of this second pilot study was to further inform the final study design and the focus of this pilot was the feedback intervention. The aim of this study was to explore how the different models of feedback (information transfer and dialogic) can be orchestrated and articulated and to gain a better understanding of how the two, different feedback models would be expressed and differentiated in a study setting. This second pilot study built on the practical experience gained in the first pilot in relation to running an experimental study. It also allowed opportunity to test an experimental study protocol on a smaller scale and ensure that the materials designed were as robust as possible for the final study. Forming specific questions helped me to clarify exactly what information and knowledge I wanted to gain from this pilot:

- Could I construct a task-specific checklist for delivery of information-transfer feedback?
- What would an integrated model of dialogic feedback and encouraged self-regulation look like in practice?
- Could the use of process goals help link the cycles of task-specific self-regulation?
- Does use of the constructed forms increase the accuracy and integrity of the feedback type?
- Would participants be able to engage with dialogic feedback?
- Could depth of feedback content be standardised between the two groups?
- Could time taken to give or construct feedback be standardised between groups?

6.1.2 Ethical approval

Ethical approval for this study was sought prior to commencement of the pilot studies. The University of Dundee Research Ethics Committee (Application 14134) granted permission on 29th October 2014. (Appendix A)

6.2 METHODS

6.2.1 Study design

6.2.1.1 Participants, recruitment and randomisation

Third year undergraduate medical students at the University of Dundee were recruited to this pilot via an invitational email to their year group. In return for participating in the pilot study, the students were offered a certificate of participation for their undergraduate portfolio. This pilot study took place over two weeks between 3rd – 14th November 2014. Recruiting medical students to a medical education study ensured the validity of the participant group. Twenty medical students were recruited in total. The number recruited was based upon the practical factors of duration of the study (two weeks) and availability and willingness of participants. Each participant was assigned a participant number (1-20) and allocated a random number via the 'random number' function in Microsoft Excel. The participants were then ranked in order of their allocated random number. Participants 1-10 were allocated to Group A (information transfer feedback) and participants 11-20 to Group B (dialogic feedback). The file depicting this randomisation process can be found in the accompanying data disk ('Pilot study 2 Participant randomisation').

Six of the 20 recruited participants failed to start the study as no mutually convenient time could be established. One further participant failed to attend the second study visit. The final participant group included seven randomised to Group A (n=7) and six randomised to Group B (n=6).

6.2.1.2 Study setting

As with the first pilot, the study was based at the Cuschieri Surgical Skills Centre, Ninewells Hospital & Medical School. However, rather than use the laparoscopic simulation room, one of the LapSim® models was moved to a separate room within the unit. This overcame the issue of room access and lack of availability encountered in the previous pilot. This room could be accessed at any time between 9am and 5pm and allowed greater flexibility when arranging participant visits.

6.2.1.3 Study protocol

Figures 33 and 34 refer to the study protocols employed in visit one and two respectively in the second pilot study. Each study visit was structured as per the 'session components' described.

Figure 33: Pilot study two, visit one protocol

Session component	**Participant group**	
	Group A; information transfer feedback (n = 7)	*Group B; dialogic feedback with self-regulation (n = 6)*
Orientation	Study information sheet issued; Consent form issued / signed; Pre-study data collection sheet completed	
Explanation of task	Participant view task instruction booklet. Opportunity for participant questions.	
Task practice	5 minutes of supervised and structured task practice Tutor giving instruction on how to use LapSim controls, aims of the task, clip applying technique but NOT tactics or technique correction	
Task performance	**Task 1 performance 1 (P1)**	
Explanation of feedback	- Explanation of information transfer feedback - Orientation to feedback checklist	- Explanation of self-regulation, dialogic feedback - Orientation to feedback form
Feedback session 1	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Task performance	**Task 1 performance 2 (P2)**	
Feedback session 2	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Task performance	**Task 1 performance 3 P3)**	
Feedback session 3	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development
Task performance	**Task 2 performance 1 (P1)**	
Debrief	Thank for time; arrange follow up visit in approximately 7 days	

Figure 34: Pilot study two, visit two protocol

Session component	**Participant group**	
	Group A; information transfer feedback (n = 7)	*Group B; dialogic feedback with self-regulation (n = 6)*
Task practice	5 minutes of supervised task practice Tutor able to remind participant how to use LapSim controls, aims of the task, clip applying technique but NOT tactics or technique correction	
Task performance	**Task 1 performance 4 (P4)**	
Feedback session 4	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Task performance	**Task 1 performance 5 (P5)**	
Feedback session 5	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development
Task performance	**Task 2 performance 2 (P2)**	
Debrief	Thank for time; answer any participant questions Collect post-study data	

In this study, the participants assume the role of the learner and the researcher the role of the tutor. All study visits and feedback sessions were conducted by myself, the lead researcher.

6.2.1.4 Study visits

A review of the medical education feedback literature shows variation in the design of experimental studies (Chapter 2, section 2.3.2). Half of the studies featured in this structured literature review were single-visit studies, with the effect of feedback measured during, and limited to, only one episode of contact with the participant. Others employed a two-visit design, in which two separate visits were required in order to capture the post-intervention data. The gap between the first and second visit varied from only 24 hours between study visits (Farquharson, 2013) and a full academic year (Wojcikowski, 2013). In this pilot study, participants were asked to complete the two tasks in two study visits, approximately one week apart. With little consensus in the medical education literature, this inter-visit gap was based on the clinical experience of operating lists often being weekly occurrences.

Due to restrictions of availability, participants completed the two visits with gaps of between four and nine days between visits (Table 20). The mean gap between study visits was 7.2 days, the median and modal was 7.5 days.

Table 20: Variation and frequency of gap (in days) between study visit 1 & 2

Number of days	Frequency (No. of participants)
4	*1*
5	*0*
6	*3*
7	*3*
8	*3*
9	*3*

6.2.1.5 Orientation

At the start of visit one, participants were provided with a study information sheet (Appendix I) and informed consent (Appendix J) was obtained. Participant information was also collected (Appendix K). This included age, sex, dominant hand, self-rated confidence level, self-rated experience level, description of experience and participant confidence in tutor (researcher). The latter was an attempt made to quantify any potential confounding factor in relation to difference in latent confidence in the tutor (researcher) that might bias results. That the latter was not blinded in this pilot study might encourage inaccuracy in the reported confidence and causes limitation in its value.

6.2.1.6 Participant task instructions

Participant instructions were given via a printed instruction booklet (Appendix L). This 22-page booklet described both tasks in detail via text and pictures taken during the various stages of task completion. Participants were asked to read this instruction booklet, opportunity for clarification of any material within the booklet was offered prior to the task one practice when participants completed a set of specific actions: grasping a vessel, changing instruments, closing a clip, and using the suction. The purpose of this method of imparting task instruction followed by a structured pre-performance practice was to standardise the baseline understanding between participants, reducing pre-study variation and promoting quality of the study.

6.2.1.7 Task practice

Each participant was given an opportunity to practice task one at the start of the study. Practice time was limited to approximately five minutes. The aim of the brief and time-controlled practice in relation to task one at the start of each session was to standardise participant understanding at the start of the study and, therefore, remove the effect of variation in understanding as a reason for variation in performance at the start of the study. By doing so, subsequent variation between performance one and two is likely to be due to motor skill and greater metacognitive engagement in task completion, rather than simply due to a better understanding of how to use the instruments, which would have been necessary to address during the first feedback session.

During task one practice, participants were supervised and asked to complete five actions: changing instrument type, grasping and moving the vessel, attempted application of a clip, rupture of the vessel via excessive tension, and use of the suction device. This practice replaced the researcher practical demonstration featured in the first pilot study and provided a better test of participant understanding. Asking participants to observe these functions assumed but did not test knowledge.
However, asking the participants to complete these five actions allowed demonstration of all of the basic skills required for task completion and ensured a basic level of understanding of these instructions. This promoted an even baseline of participant competence prior to data collection and, thus, promoted the quality of the study.

There was no practice in relation to task two. The purpose of this task was to assess the participant's ability to apply the simulated laparoscopic motor and metacognitive skills acquired during task one performance and feedback to another task. Therefore, the very first performance of that task was of interest. For that reason, the only instructions given were provided via the instruction booklet (Appendix L) although the participants were given an opportunity to ask questions and clarify understanding of task two at the beginning of visit one.

6.2.1.8 Task performance and repetition

Task one (vessel ligation and division) was performed three times during visit one (performances 1, 2 and 3) and a further twice during visit two (performance 4 and 5). A feedback session followed each task one performance. The increased number of repetitions of performance allowed for more detailed examination of the changes and trends in performance over the course of the study. Comparison of performances 2 to 1 and 3 to 2 in visit one, and 5 to 4 in visit two, quantified the immediate effect of the feedback sessions (intra-visit feedback effects). Comparison of performance 3 and 4 illustrated the effect of a hiatus on participant performance of this task (inter-visit feedback effects).

Task two (vessel sectioning) was performed at the end of visits one (performance 1) and two (performance 2). Analysis of performance 1 and 2 separately, across the participant groups, conferred the ability of participants to transfer simulated laparoscopic skills from one task to another in visit one and two respectively. Direct comparison of these two performances investigates the effect of a hiatus on participant performance of this task (inter-visit feedback effects).

6.2.1.9 Study debrief

At the end of visit two, participants were asked to complete the post study data collection form (Appendix M). This form asked participants to rate their post-study self confidence in performing laparoscopic tasks and participant confidence in the tutor (researcher) in providing useful feedback. The data collected was analysed and correlation between feedback type and participant self-confidence and tutor (research) confidence were investigated. Participants were asked to rate their overall satisfaction in the feedback provided as it was thought that inter-group variation in these ratings may signify a source of bias. Finally, participants were thanked for participating in the study.

6.2.2 The psychomotor task

6.2.2.1 Task one: Clip ligation and division

Based up on the experience from the first pilot study, the 'clip applying' task on the LapSim® model was chosen again for the second pilot. This task is described in detail in the previous chapter (Chapter 5, section 4.3). In brief, the aim of the simulated task is to ligate a vessel lying in the abdominal cavity before then dividing the vessel. The vessel is ligated by placing a correctly applied clip in each of the green target areas (one on the left and one on the right side of the vessel) and the vessel is then divided in the blue cutting target area (in the centre of the vessel) using the scissors. The task will automatically end once completed.

6.2.2.1.1 Task one settings

Box 7 shows the settings selected for this pilot study. These settings were unchanged from the first pilot study.

Box 7: Pilot two, task one LapSim® 'Clip application' task settings

Camera angle	-20 degrees (vessel slightly oblique to viewer, left side closer)
Clip target area size (mm)	4
Cutting target area size (mm)	4
Stretch sensitivity	Low
Spontaneous bleeding	Off
Bleeding flow rate (L/min)	1

6.2.2.1.2 Task one measures of performance

The same quantitative measures of task performance were used in the second pilot study as the first. The LapSim® collects eight separate measures of performance in relation to the clip applying task as detailed in Box 8. These measures assess participant efficiency and accuracy.

Box 8: Pilot two, task one LapSim® 'Clip application' measures of performance	
Total time to complete task	*(TTC, in seconds)*
Right instrument path length	*(RIPL, in centimetres)*
Left instrument path length	*(LIPL, in centimetres)*
Number of incomplete (missed) target areas	*(ITA, raw number, range 0-3)*
Number of badly placed clips	*(BPC, raw number)*
Number of dropped clips	*(DC, raw number)*
Maximum blood vessel strength	*(BVS, expressed as percentage of maximum tolerated)*
Blood loss	*(BL, in millilitres)*

Total time to complete task (TTC) and the sum of right and left instrument path lengths (combined instrument path length, CIPL) are measures of efficiency. Number of badly placed clips (BPC) is a measure of accuracy. The time taken to complete a task is a widely utilised and accepted measure in the medical educational literature (Pusic et al, 2014) and features in several of the medical education experimental studies investigating feedback (Rogers et al, 2000; Judkins et al, 2006; Xeroulis et al, 2007; Grancharov et al, 2007; O'Connor et al, 2008; Kruglikova et al, 2010; Boyle et al, 2011; Kannappan et al, 2012; Strandbygaard et al, 2013; Paschold et al, 2014). Instrument path length measures economy of movement; this is a widely discussed and accepted marker of surgical efficiency and competence in clinical practice and has been used previously in in the medical education literature (Judkins et al, 2006; Grancharov et al, 2007; O'Connor et al, 2008; Boyle et al, 2011; Paschold et al, 2014). The number of clips not placed correctly is a measure of how well a participant is able to understand, assimilate, apply and repeat the specific technique required for this subtask. This, therefore, is a marker of accuracy and quality of performance. In their experimental feedback study, which featured the LapSim® clip applying task, Paschold et al (2014) incorporated this measure in the 'number of errors' measure of accuracy of task performance.

6.2.2.2 Task two: Vessel sectioning

In this pilot study, a second task was incorporated as a 'cross-over' task. The participant's completion of this task, which requires some similar basic laparoscopic skills as featured in task one, was used as a measure of skill transferability. The LapSim® 'vessel sectioning' task was chosen for task two.

6.2.2.2.1 Task technique

The aim of this task is to cut the simulated vessel into sections and deposit the cut sections into a target zone. For this task, the participant has two fixed instruments: a grasper (right hand) and a pair of cautery scissors (left hand), which heat up and section the vessel when the jaws are closed around the vessel and the foot pedal is depressed. The vessel can only be grasped in the green grasping zone (at the free end of the vessel, Figure 35) and cut in the blue cutting zone (adjacent to the green grasping zone, Figure 36). This blue zone will only appear if the vessel is stretched and put under adequate tension. Once obtained, the cut section of vessel is then placed in the red hemisphere (Figure 37), which will turn green to indicate correct positioning (Figure 38) and the grasp released (Figure 39). The cycle begins again and the vessel length shortens after each section is removed. The participant must attempt to section the vessel five times and the task will end automatically after the fifth attempt. The pitfalls in this task are that the vessel cannot be cut outside the blue zone, the cut section of vessel can be dropped (failure to deposit in the red hemisphere results in a 'drop failure') and that the vessel will rupture if placed under too much tension (Figures 40 and 41).

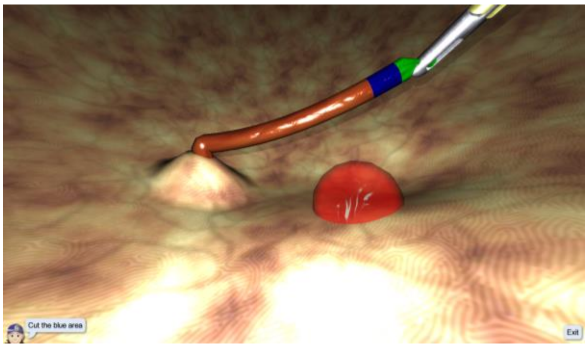


Figure 35: The vessel is held by the graspers in the green zone, placed under tension and thus the blue cutting zone is visible.

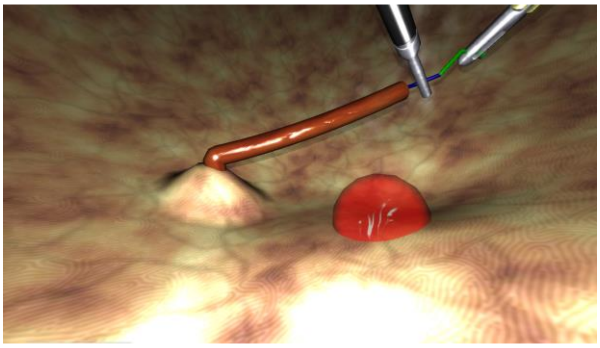


Figure 36: The jaws of the cautery scissors are closed around the vessel in the blue cutting zone. The foot pedal is depressed to activate the scissors.

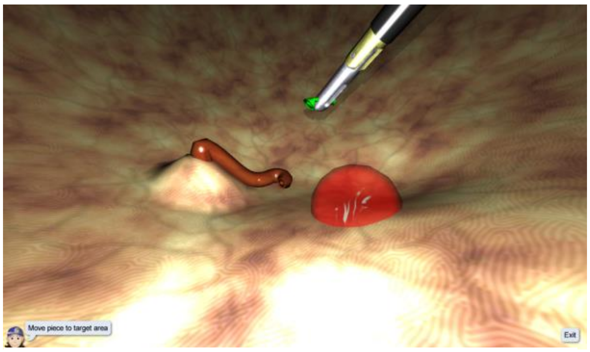


Figure 37: The vessel section is held by the graspers and can be transferred and held inside the red hemisphere.

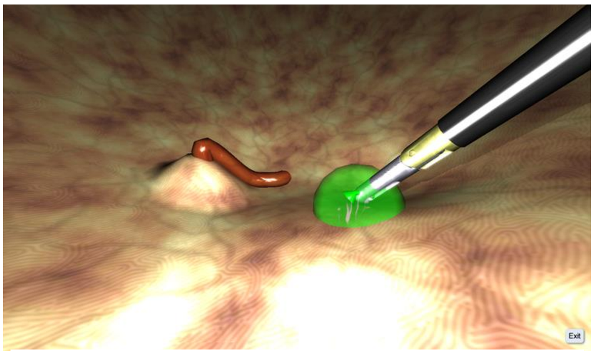


Figure 38: The red hemisphere turns green to show the section is held within its boundaries.

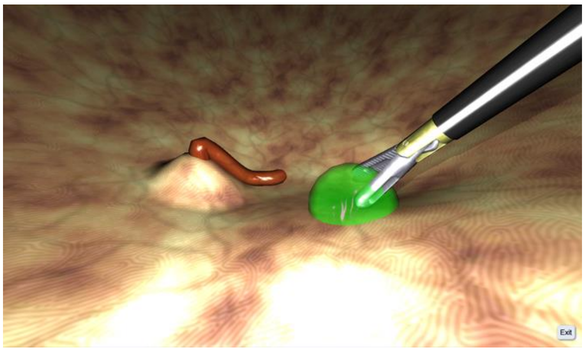


Figure 39: The vessel section is released and the cycle restarts and continues until 5 attempts at sectioning the vessel have been completed.

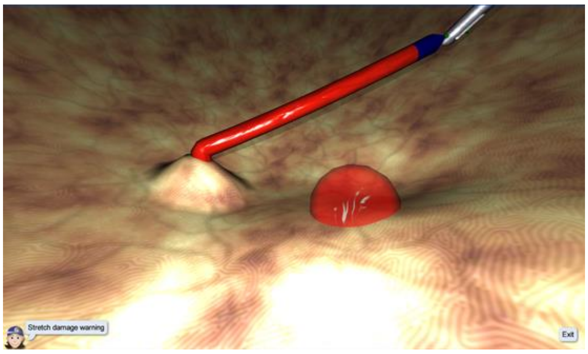


Figure 40: The vessel is under significant tension as shown by the bright red colour.

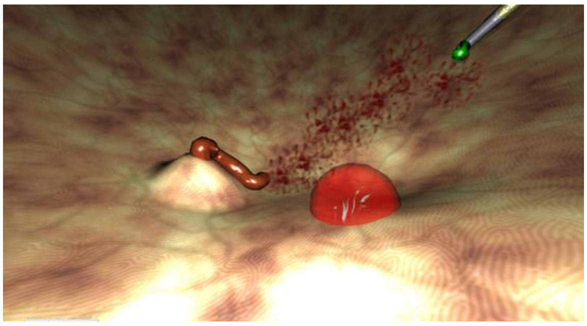


Figure 41: The vessel has ruptured due to excessive tension. The cycle restarts and continues until 5 attempts at sectioning the vessel have been completed.

6.2.2.2.2 Task two setting options

The LapSim® software allows configuration of the task, which allows reproducibility and manipulation of difficulty level. Configurations fall into two broad categories:

Camera options:	At 0° the camera faces the vessel straight on. By entering either negative or positive values, the camera can be positioned to look at the vessel either from the left or right. A moving or non-moving camera can be selected.
Vessel options:	The size of both clipping and cutting target areas (small, medium or large) can be set, as can the stretch sensitivity of the vessel (low, medium or high), which dictates how easily it is ruptured.

6.2.2.2.3 Task two selected task settings

Box 9 shows the settings selected for the purposes of this pilot study. These settings were selected as a starting point based on both the programme 'automatic' settings and informal investigator trials.

Box 9: Pilot two, task two LapSim® 'Clip application' task settings

Camera angle	-10 degrees (vessel slightly oblique to viewer, left side closer)
Camera type	Non-moving
Clip target area size	Medium
Cutting target area size	Medium
Stretch sensitivity	Low

6.2.2.2.4 Task two measures of performance

The LapSim® collects eight separate measures of performance in relation to the clip applying task as detailed in Box 10. These measures can be used to assess participant efficiency and accuracy.

Box 10: Pilot two task two LapSim® 'Clip application' measures of performance	
Total time taken	*(TTC, in seconds)*
Right instrument path length	*(RIPL, in centimetres)*
Left instrument path length	*(LIPL, in centimetres)*
Number of sections obtained	*(NoS, raw number, range 0-5)*
Vessel rupture rate	*(VRR, %)*
Number of dropped sections	*(NoDS, raw number, range 0-5)*
Maximum blood vessel stretch damage	*(BVS, expressed as percentage of maximum tolerated)*
Frequency of abdominal wall damage	*(ABD, raw number)*
Maximum depth of abdominal wall damage	*(MDD, millimetres)*

6.2.3 The feedback intervention

Feedback forms for each of the feedback group were based upon the characteristics of information transfer and dialogic feedback as described in chapter 3 (Section 3.6, Box 4) and the experience gained during the first pilot study.

6.2.3.1 Information transfer feedback

The revised information transfer feedback form is seen below in Figure 42 and is attached as Appendix N. This form retains the global and task specific elements from the original form as these were useful in providing comprehensive and detailed feedback. Within the space for free text comments relating to each of the elements, notes relating to common and recurrent observations associated with each element were pre-entered to shortcut researcher note taking during task performance observation. During the task performance, the researcher highlighted the pre-entered comments relevant to that performance; one highlighter colour (green) was code for 'positive' aspects of the performance (which was discussed in the subsequent feedback session and reinforced) and a second colour (pink) coded for 'negative' aspects of the performance (which was discussed in the subsequent feedback session and suggestions given for how to improve this aspect of performance). Any other comments relating to aspects of the performance not included in the pre-entered comment were written in by hand and highlighted in the appropriate colour.

During Group A feedback sessions, the information transmission feedback form was completed in full. The researcher/tutor worked through each of the global and task specific components, providing the participant with a rating for each component (N, D or S) and discussed their observational comments relevant to each task component. The feedback session concluded with the tutor providing an overall rating of the performance via the global rating (0-4). These feedback sessions were highly structured and the discussion was tutor-led.

Date:

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education
Pilot study: Group A
Feedback checklist

Participant number: Task #: Level: Time for FB:

	Component	Rating			Comments
G1	Follows and agreed, logical sequence or protocol for the procedure	N	D	S	> At start > Adaptation > At end – bleeding / clip removal > Problem solving
G2	Consistently handles tissue well with minimal damage	N	D	S	> Arm position > Danger to vessel when 2 instruments being used > Inadvertent movements > Ruptured vessel
G3	Controls bleeding promptly	N	D	S	> Is 1st thought suction > Clipping held vessel before transferring to 2nd bleeder > Use of grasper for temporary arrest > Recognising bleeding vessel
G4	Uses instruments appropriately and safely	N	D	S	> Instrument selection > Dexterity > Instruments always in view > Orientation > Co-ordination of instruments
G5	Proceeds at an appropriate pace with economy of movement	N	D	S	> Bleeding - complete blood removal before changing task > Check clip good before releasing vessel > Multitasking with 2 instruments > Reducing instrument travel > 2nd hand neglect
G6	Deals calmly and effectively with unexpected events and complications	N	D	S	> Emotional reaction > Tactical reaction
T1	Grasps vessel carefully with grasper to allow clip application	N	D	S	> Position of grasper on vessel > Positioning of vessel > Holding vessel steady
T2	Applies clips correctly to either side of vessel	N	D	S	> Position of clip - Within target areas - To stop bleeding - Taught not flaccid vessel > Maximising clip efficacy / minimising BPC
T3	Cuts vessel safely with scissors	N	D	S	
T4	Retrieves dropped clips	N	D	S	> Use of 2 graspers > Use of closest grasper > Dexterity to reduce accidental clip closure

Level of performed task		
0	Unable to complete task	
1	Able to complete task with significant difficulty in several components	
2	Able to complete task with mild-moderate difficulty in several components	
3	Able to perform all elements of task with mild difficulty in only some components	
4	Able to perform all elements of task, exhibiting high skill level	

Figure 42: Information transfer feedback form

6.2.3.2 Dialogic feedback

The revised dialogic feedback form (based on the pilot study one version) is seen below in Figure 43 and is attached as Appendix O. This form retained the global and task specific elements from the original form as these were successful in organising comprehensive and detailed feedback for discussion during the dialogic feedback sessions. In a similar fashion to the information transfer form, notes relating to common and recurrent observations associated with each task element were pre-entered, shortcutting researcher note taking during task performance observation and facilitating more detailed observation. During the task performance, the researcher highlighted the pre-entered comments relevant to that performance; one highlighter colour (green) relating to 'positive' aspects of the performance (which may have been discussed in the subsequent feedback session and reinforced) and a different colour (pink) for 'negative' aspects of the performance (which may have been discussed in the subsequent feedback session). Any other comments that related to aspects of the performance not included in the pre-entered comments, were written in by hand and highlighted in the appropriate colour.

During the explorative discussions featured in Group B feedback sessions, the dialogic feedback form was used as a researcher aide-mémoire for contributing detailed and specific observations to help participant understanding and self-analysis. They were also used to help the researcher direct or provide starting points for feedback sessions when participants are unable to do this. The dialogic feedback sessions were by design less prescriptive than the information transmission sessions; the form itself was not directly referred to or 'worked through' in the same way. The purpose of the dialogic feedback form is to promote discussion of process rather than record outcomes and judgement.

6.2.3.2.1 The use of process goals

The encouraged use of process goals was incorporated into the revised dialogic feedback form to promote the feed-forward self-regulatory element of the feedback sessions. Previous studies examining the intra-task self-regulatory process have highlighted the importance of process goals in successful task performance (Cleary, 2011; Artino, 2014). Other interventional studies, which aimed to investigate the effect of encouraged self-regulation of learning have failed to promote the use of process goals (Brydges, 2012; Shanks, 2013). In this pilot study, participants in the dialogic feedback group were asked for two process goals at the end of each feedback session. These process goals should relate to two behaviours that they would like to use in their next task performance. Reflection on these goals formed part of the subsequent feedback session. Participants in this group were reminded of the process goals they set at the end of the first study visit (post performance 3 of task one) at the beginning of the second visit (prior performance 4 of task one). The use of process goals in this way provided a direct link between the reflection phase of the previous performance and the planning phase of the subsequent performance, facilitating the cycle of self-regulation.

Date:

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study: Group B
Feedback notes

Participant number: Task #: Level: Time taken for FB:

	Component	**Systematic / tactical awareness**	**Technical skills**
G1	Logical sequencing	> At start > Adaptation > At end – bleeding / clip removal > Problem solving	
G2	Tissue handling		> Arm position > Danger to vessel when 2 instruments being used > Inadvertent movements > Ruptured vessel
G3	Use of instruments	> Instrument selection	> Dexterity > Instruments always in view > Orientation > Co-ordination of instruments
G4	Pace with economy of movement	> Bleeding - complete blood removal before changing task > Check clip good before releasing vessel	> Multitasking with 2 instruments > Reducing instrument travel > 2nd hand neglect
G5	Deals calmly and effectively with unexpected events and complications	> Emotional reaction > Tactical reaction	
G6	Awareness of self-regulation	> Tasks within tasks > Ensures clips on before releasing / cutting > Evidence of checking work	
T1	Use of grasper to allow clip application	> Position of grasper on vessel > Positioning of vessel	> Holding vessel steady, reducing
T2	Applies clips correctly	> Position of clip - Within target areas - To stop bleeding - Taught not flaccid vessel	> Maximising clip efficacy / minimising BPC > "Sneaking up"
T3	Control of bleeding	> Is 1st thought suction > Clipping held vessel before transferring to 2nd bleeder > Use of grasper for temporary arrest	> Recognising bleeding vessel
T4	Minimises dropped clips / Optimises method of retrieval	> Use of 2 graspers > Use of closest grasper > Awareness of number dropped	> Dexterity to reduce accidental clip closure

GOALS: **1.**

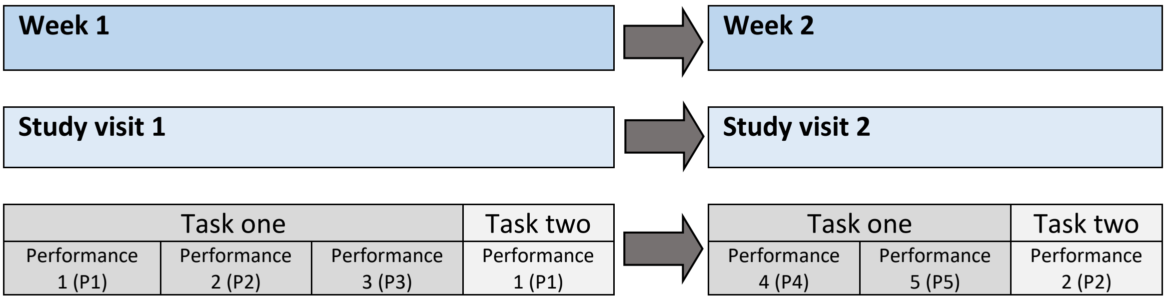
2.

Figure 43: Dialogic feedback form

6.3 RESULTS

The raw data and tabulated results and analysis for pilot study two are available for viewing in the accompanying ‘Data disk’ (filename ‘Pilot study 2 raw data’ and ‘Pilot study 2 results’ respectively). For the purposes of clarity, the two study groups will be referred to as A (information transfer feedback, ITF) and B (dialogic feedback, DF) throughout this results section. Figure 44 below illustrates the timing of visits and task performances.

Figure 44: Chart illustrating timing of study visits and task one and task two performances (P1 – P5 and P1 – P2 respectively)



6.3.1 Demographics

Group A (ITF) and B (DF) were similar with respect to participant handedness. There was an inequality with respect to sex (Table 20). The relevance of this is unknown.

Table 20: Demographic analysis of participants

Group	***Sex***		***Handedness***	
	Female	*Male*	*Right*	*Left*
A (ITF)	*3*	*4*	*6*	*1*
B (DF)	*4*	*2*	*6*	*0*

The age of participants of group A and B was investigated via an independent Student t-test and no significant difference between Group A and B was found (Table 21).

Table 21: Age analysis of participants

		Group A (ITF)	Group B (DF)	P values	
Age (years)	Mean (SD)	22.57 (3.51)	21.83 (3.06)	T-test	0.693

6.3.2 Analysis of practice time

The time participants of Group A (ITF) and B (DF) spent practicing prior to performance one (P1) was investigated via an independent Student t-test and no significant difference between Group A and B was found (Table 22), although the difference did approach significance (p= 0.069). The data shows that on average Group B spent less time practicing than Group A. Table 23 illustrates the analysis of participant practice time in visit two, prior to performance 4 (P4). This shows that on average both groups spent less time practicing in visit two compared to visit one and that there was no significant difference in the time spent between the two groups (p=0.880).

Table 22: Analysis of time spent practicing (Visit one)

		Group A (ITF)	Group B (DF)		P values	
Time spent practicing (seconds)	Mean (SD)	321.43 (32.75)	289.17 (24.85)		T-test F-test	0.069 0.561

Table 23: Analysis of time spent practicing (Visit two)

		Group A (ITF)	Group B (DF)	P values	
Time spent practicing (seconds)	Mean (SD)	150.67 (33.95)	139.67 (89.28)	T-test F-test	0.880 0.830

6.3.3 Analysis of feedback volume

The volume of feedback given to participants in each group (length of feedback sessions) was controlled in order to limit this as a confounding factor between participants and the study groups and promote the quality of the study design. All feedback sessions were timed and an effort was made to limit time to five minutes.

There was no significant difference in the volume of feedback provided to either group (Table 24, p=0.342). There was reduced variation in feedback session length in Group A, which is consistent with the information transfer feedback being more structured, but the difference was not statistically significant (p=0.142).

Table 24: Analysis of length of feedback sessions (session 1 – 5)

		Group A (ITF)	Group B (DF)	P values	
Length of feedback session (seconds)	Mean (SD)	321.51 (66.03)	302.83 (85.99)	T-test F-test	0.342 0.142

6.3.4 Summary of results

No significant differences in relation to the pre-study experience, pre-study confidence or demographics were found in relation to the study groups. This supports the random allocation method used in this study.

Inter-group comparisons in relation to markers of efficiency (Time Taken to Complete task, TTC, and Combined Instrument Path Length, CIPL) showed no statistically significant differences in performance over the whole study (Performances 1-5). However, analysis of both measures resulted in the same inter-visit trend being identified. At the start of visit two (Performance 4), Group B (dialogic feedback) were more efficient than they had been at the both the start (Performance 1) and the end (performance 3) of visit one. In contrast, Group A (information transfer feedback) became less efficient at the start of visit two in comparison to both performance 1 and 3.

Inter-group comparisons in relation to accuracy (number of Badly Placed Clips, BPC) highlighted interesting differences and trends in performance. On average, Group A became less accurate over the course of the study, with a higher average number of BPC in Performance 5 than Performance 1. In contrast, Group B became more accurate. The same inter-visit trend was observed in relation to accuracy as was seen with efficiency; Group B performance at the start of visit two (Performance 4) was superior in comparison to Group B Performance 1 or 3. However, Group A performance at the start of visit two was less accurate than these previous performances. In addition, there was a statistically significant difference identified in relation to the reliability of improvement during visit one. Intra-visit analysis of visit one (Performances 1-3) showed less variation in the rate of improvement of participants in Group B in comparison to those in group A (p=0.009).

Inter-group performance analysis specifically focused on Performance 4 (at the start of visit two) highlighted statistically significant differences in the intra-group variation in relation to measures of efficiency (TTC, CIPL) and accuracy (BPC). Group B performance was much more consistent at this point in the study in comparison to Group A participants (p≤0.002). The inter-group differences in efficiency of this performance (TTC, CIPL) approached significance (TTC p=0.096, CIPL p=0.085), with Group B on average more efficient than Group A. The same trend was observed in relation to accuracy but statistical analysis did not suggest a significant difference (p=0.198).

Analysis and inter-group comparisons of efficiency of performance (TTC) in relation to the cross-over task (Task 2) did not show any statistically significant differences between the two groups in either Performance 1, Performance 2 or the gradient of change between Performance 1 and 2. However, a vastly reduced variation in this measure of performance was observed in Group B compared to Group A in relation to Performance 2 (visit two) and the gradient of change between Performance 1 and 2. This again suggests that Group B performance was more reliable and predictable.

Statistical analysis of post-study participant self-confidence and confidence in tutor feedback did not show any significant differences between the groups. Both groups reported very high post-study confidence in tutor feedback.

The quantitative analysis of this pilot study was more successful in identifying interesting trends as opposed to robust statistical differences. The reason for this is either that no true difference exists or that the study is underpowered and fails to illustrate differences that do exist (Type I error) due to the small number of participants and the large variation in performance measures.

6.3.5 Pre-study participant confidence & experience

The pre-study participant confidence in self in relation to performing laparoscopic tasks and in the tutor (researcher) in providing useful feedback in relation to laparoscopic tasks was measured at the start of this study as differences may have contributed a confounding factor. Participants were asked to score each of these factors between 1 (low) and 5 (high). Potential differences between Groups A and B were investigated via an independent Student t-test and no significant differences were found (Table 25, p=0.80 and 0.55 respectively). The results illustrate a relatively low pre-study self-confidence (Group A (ITF) mean = 2.00 and Group B (DF) mean = 2.17) but high pre-study confidence in tutor (researcher) amongst participants (Group A (ITF) mean = 4.17 and Group B (DF) mean = 4.33).

Table 25: Pre-study confidence and experience analysis of participants

		Group A (ITF)	Group B (DF)		P value
Pre-study self-confidence (range 1-5)	Mean (SD)	2.00 (1.15)	2.17 (1.17)	T-test	0.801
Pre-study tutor-confidence (range 1-5)	Mean (SD)	4.17 (0.41)	4.33 (0.52)	T-test	0.551

Participants self-reported pre-study experience levels were low (Group A (ITF) mean = 1.14 and Group B (DF) mean = 1.17), which was supported by the free-text descriptions of experience. A potential difference between Group A and B were investigated via an independent Student t-test and no significant differences were found (Table 26, p=0.916).

Table 26: Analysis of participant pre-study experience rating

		Group A (ITF)	Group B (DF)		P value
Pre-study experience rating (range 1-5)	Mean (SD)	1.14 (0.38)	1.17 (0.41)	T-test	0.916

6.3.6 Task one analysis

6.3.6.1 Task one analysis: Performance one

Quantitative analysis of performance one was performed to examine for potential differences in baseline ability between Group A and B. The results are summarised in Table 27.

Table 27: Performance one analysis

		Group A (ITF)	Group B (DF)	Statistical comparisons	
Time taken to Complete (TTC) (seconds)	Mean Variance SD	227.33 12836.67 113.30	255.00 17394.40 131.89	T-test	0.705
Combined instrument path length (CIPL)	Mean Variance SD	5.91 16.54 4.07	6.79 20.97 4.58	T-test	0.731
Number of badly Placed clips (BPC)	Mean Variance SD	1.00 1.20 1.10	3.67 17.87 4.23	T-test	0.185

The P1 quantitative measures of performance achieved by participants in Groups A and B was investigated via an independent Student t-test and no significant difference between Group A and B was found in either the measures or efficiency (time taken to complete and combined instrument path length) or accuracy (number of badly placed clips).

This would suggest that there was no difference in the baseline abilities between the two study groups.

6.3.6.2 Task one analysis: Time taken to complete (TTC)

The time taken to complete the task (TTC) for each performance (P1 – P5) was recorded for each participant. Figure 44 illustrates the relative timing of each task one performance. TTC is a measure of efficiency of task completion. The trend of each participant's performance during the course of the study was examined by calculating the gradient of the slope between data points of interest. Change in performance (improvement or deterioration) of Group A (ITF) and B (DF) was investigated and compared via an independent Student t-test using the gradient of the line between data points. Variation in group performance (how consistent each group was) was investigated and compared via an F-test, comparing the variation in gradients of the lines between data points.

When comparing the difference between performances of the task, a negative gradient shows TTC has reduced between the two data points, that is, efficiency has improved. Conversely, a positive gradient shows TTC has increased between the two data points, that is, efficiency has deteriorated.

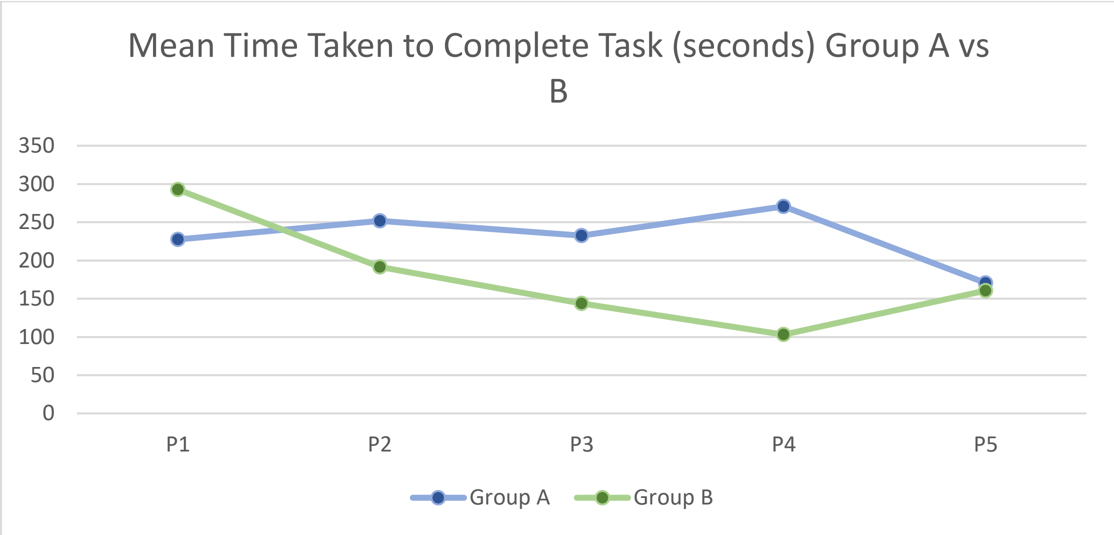
6.3.6.2.1 Performance 1 – 5 (whole study length)

There was no significant difference between Groups A (ITF) and B (DF) in the change in efficiency of performance (as measured using TTC) between P1 and P5 (Table 28, p=0.563) nor in the consistency of this change (p=0.872). The average performance of both groups improved between the beginning (P1) and the end (P5) of the study (Figure 45).

Table 28: Analysis of TTC change in performance (P1 – P5) (seconds)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	-9.44	-27.20	T-test	0.563
Variance	2444.09	2842.63	F-test	0.872
SD	49.43	53.32		

Figure 45: Mean TTC for P1 – P5, Group A (ITF) vs B (DF)



6.3.6.2.2 Performance 1 – 3 & 4 – 5 (intra-visit change)

Examining the change in efficiency of performance between P1 and P3, and P4 and P5, examines the change in efficiency during visit one and two respectively; the intra-visit effect of feedback. This examines the immediate effect of task repetition and feedback.

No significant difference between Groups A and B were identified in relation to intra-visit change in efficiency of performance (as measured using TTC) nor in the consistency of this change (Tables 28 and 29). However, an interesting difference in trends can be observed. During visit one (Table 28), participants in Group B (DF) tended to improve in efficiency, whereas the performance of those in Group A (ITF) was fairly static (Figure 45). This trend is reversed in visit two (Table 29); in visit two participants in Group A (ITF) improved in efficiency on the whole (gradient -100.04) but those in Group B (DF) deteriorated (gradient +68.58).

Table 28: Analysis of TTC change in performance visit one (P1 – P3) (seconds)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	+2.50	-55.58	T-test	0.335
Variance	7040.40	12492.94	F-test	0.544
SD	83.91	111.77		

Table 29: Analysis of TTC change in performance visit two (P4 – P5) (seconds)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	-100.04	+68.58	T-test	0.130
Variance	32442.20	30318.30	F-test	0.943
SD	180.12	174.12		

6.3.6.2.3 Performance 1 – 4 & 3 – 4 (inter-visit change)

Examining the change in efficiency of performance between P1 and P4, and P3 and P4, examines the change in efficiency between visit one and two respectively; the inter-visit effect of feedback. This examines the effect of a hiatus of one week between feedback and task performance. Looking at the change between P1 and P4 compares the performance at the start of visit two to that at the start of visit one Table 30). Looking at the change between P3 and P4 compares the performance at start of visit two to that at the end of visit one (Table 31).

Table 30: Analysis of TTC change in performance, visit 1 vs visit 2 (P1 – P4) (seconds)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	+11.07	-52.04	T-test	0.152
Variance	7225.01	2284.91	F-test	0.232
SD	85.00	47.80		

Table 31: Analysis of TTC change in performance, visit 1 vs visit 2 (P3 – P4) (seconds)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	+38.33	-43.63	T-test	0.276
Variance	7903.87	21606.76	F-test	0.294
SD	88.90	146.99		

No significant difference between Groups A and B were identified in relation to inter-visit change in efficiency of performance (as measured using TTC) nor in the consistency of this change (Tables 30 and 31). However, again, an interesting difference in trends can be observed. The average gradient of change between P1 and P4 and P3 and P4 in Group A was positive (+11.07 and +38.33 respectively). This illustrates that, on average, participants in Group A were less efficient at the start of visit two that they had been at either the beginning or end of visit one. This is seen in Figure 45. However, the average gradient of change between P1 and P4 and P3 and P4 in Group B was negative (-52.04 and -43.63 respectively). In fact, all participants in Group B were faster at completing the task at the start of visit two (P4) than they were at the beginning of visit one (P1).

6.3.6.3 Task one analysis: Combined instrument path length (CIPL)

The combined instrument path length (CIPL), how far the participants needed to move both instruments in order to perform the task, for each performance (P1 – P5) was recorded for each participant. CIPL, like TTC, is a measure of efficiency of task completion. Figure 44 illustrates the relative timing of each task one performance. The trend of each participant's and group performance during the course of the study was examined by calculating the gradient of the slope between data points of interest. Performance and change in performance (improvement or deterioration) of Group A and B was investigated in a similar fashion to that of TCC (independent Student t-test and f-test for variation).

A negative gradient shows CIPL has reduced between the two data points, that is, efficiency has improved. Conversely, a positive gradient shows CIPL has increased between the two data points, that is, efficiency has deteriorated.

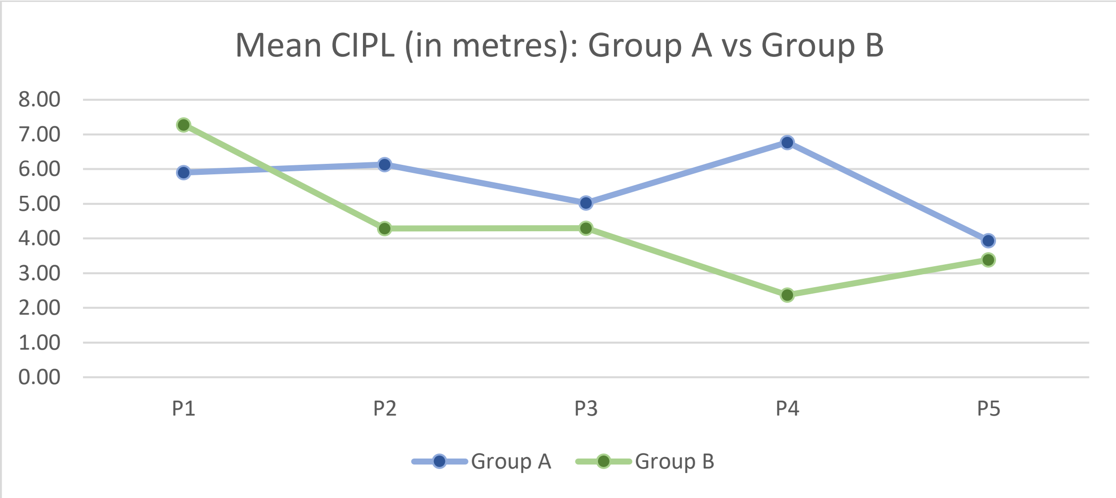
6.3.6.3.1 Performance 1 – 5 (whole study length)

Table 32 shows that there was no significant difference between Groups A and B in the change in efficiency of performance (as measured using CIPL) between P1 and P5 (p=0.553) nor in the consistency of this change (p=0.916). The pattern of change in mean efficiency measured via CIPL is illustrated in Figure 46.

Table 32: Analysis of CIPL change in performance (P1 – P5) (metres)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	-0.33	-0.86	T-test	0.533
Variance	1.96	2.16	F-test	0.916
SD	1.40	1.47		

Figure 46: Mean CIPL for P1 – P5, Group A (ITF) vs B (DF)



6.3.6.3.2 Performance 1 – 3 & 4 – 5 (intra-visit change)

No significant difference between Groups A (ITF) and B (DF) were identified in relation to intra-visit change in efficiency of performance (as measured using CIPL) nor in the consistency of this change (Tables 33 and 34). During visit one (Table 33) both groups improved, although the improvement is more marked amongst participants in Group B (Figure 46). However, an interesting difference in trend is observed in visit two (P4 and P5); one similar to the other measure of efficiency investigated (TTC). Participants in Group A (ITF) improved in efficiency on the whole (gradient -2.83) but those in Group B (DF) deteriorated (gradient +1.18). This difference approached but did not reach significance (Table 34, p=0.112).

Table 33: Analysis of CIPL change in performance (P1 – P3) (metres)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	-0.44	-1.24	T-test	0.664
Variance	4.94	14.10	F-test	0.275
SD	2.22	3.75		

Table 34: Analysis of CIPL change in performance (P4 – P5) (metres)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	-2.83	+1.18	T-test	0.112
Variance	14.61	17.03	F-test	0.870
SD	3.82	4.13		

6.3.6.3.3 Performance 1 – 4 & 3 – 4 (inter-visit change)

Exploring the change in efficiency of performance between P1 and P4, and P3 and P4, explores the change in efficiency between visit one and two respectively. This examines the effect of a hiatus between task performance and feedback.

No significant difference between Groups A and B were identified in relation to inter-visit change in efficiency of performance (as measured using CIPL) nor in the consistency of this change (Tables 35 and 36). However, again, an interesting difference in trends can be observed. The average gradient of change between P1 and P4 and P3 and P4 in Group A was positive (+0.15 and +1.75 respectively). This illustrates that, on average, participants in Group A (ITF) were less efficient at the start of visit two that they had been at either the beginning or end of visit one. This is seen in Figure 46. However, the average gradient of change between P1 and P4 and P3 and P4 in Group B was negative (-1.35 and -1.89 respectively). In fact, all but one of the participants in Group B were more economical in their movements when completing the task at the start of visit two (P4) than they were at the beginning of visit one (P1).

Table 35: Analysis of CIPL change in performance (P1 – P4) (metres)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	+0.15	-1.35	T-test	0.219
Variance	5.29	2.40	F-test	0.408
SD	2.30	1.55		

Table 36: Analysis of CIPL change in performance (P3 – P4) (metres)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	+1.75	-1.89	T-test	0.146
Variance	7.70	22.92	F-test	0.256
SD	2.77	4.79		

6.3.6.4 Task one analysis: Number of badly placed clips (BPC)

The number of badly placed clips during each performance (P1 – P5) was recorded for each participant. BPC is a measure of accuracy; it is essentially an error count. The trend of each participant's and group performance during the course of the study was examined by calculating the gradient of the slope between data points of interest. Performance and change in performance (improvement or deterioration) of Group A (ITF) and B (DF) was investigated in a similar fashion to that of TCC and CIPL (independent Student t-test and f-test for variation).

A negative gradient shows BPC has reduced between the two data points, that is, accuracy has improved. Conversely, a positive gradient shows BPC has increased between the two data points, that is, accuracy has deteriorated.

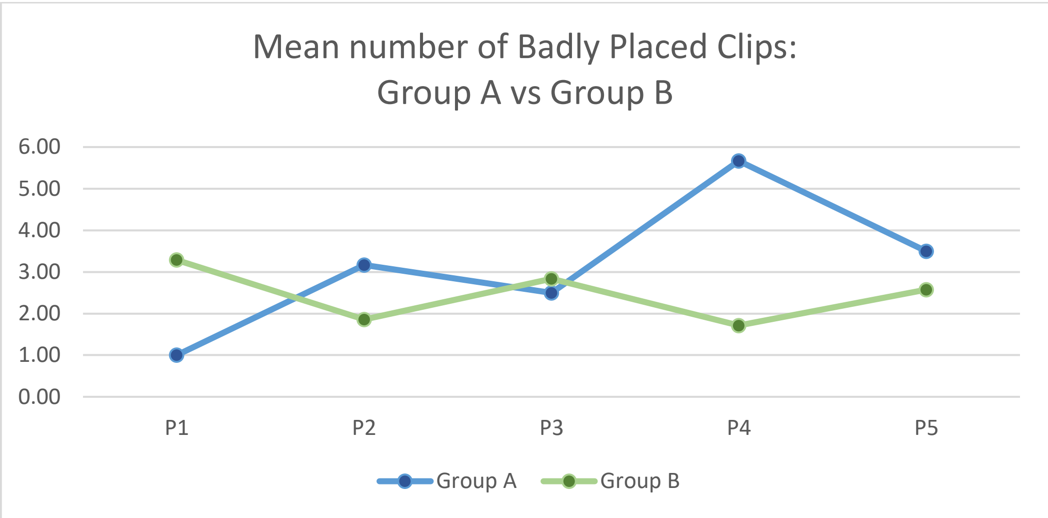
6.3.6.4.1 Performance 1 – 5 (whole study length)

Analysis showed that, on average, participants in Group A (ITF) became less accurate over the course of the study (Table 37, gradient +0.75). This was an unexpected finding. Group B (DF) participants did improve on average. However, there was no significant difference between Groups A and B identified (p>0.05), potentially due to the statistical design limitations of this study. The pattern of change in mean accuracy measured via BPC is illustrated in Figure 47.

Table 37: Analysis of BPC change in performance (P1 – P5) (raw number)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	+0.75	-0.13	T-test	0.226
Variance	1.80	1.14	F-test	0.591
SD	1.34	1.07		

Figure 47: Mean BPC for P1 – P5, Group A (ITF) vs B (DF)



6.3.6.4.2 Performance 1 – 3 & 4 – 5 (intra-visit change)

Examining the change in accuracy of performance between P1 and P3, and P4 and P5, examines the change in efficiency *during* visit one and two respectively. This examines the immediate effect of task repetition and feedback.

No significant difference between Groups A and B were identified in relation to intra-visit change in efficiency of performance as measured using BPC (Tables 38 and 39, p=0.479 and 0.167 respectively). Group A (ITF) participants tended to deteriorate in relation to this measure of accuracy during visit one but improved dramatically during visit two (gradients +0.75 and -2.17 respectively). However, whilst group B (DF) improved on average during visit one, their accuracy reduced during visit two (gradients -0.42 and +1.00 respectively). A finding of note is that Group B participants were significantly more reliable in their rate of change (improvement) than Group A during visit one (Table 38, p=0.009).

Table 38: Analysis of BPC change in performance (P1 – P3) (raw number)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	+0.75	-0.42	T-test	0.479
Variance	0.88	13.44	F-test	**0.009**
SD	0.94	3.67		

Table 39: Analysis of BPC change in performance (P4 – P5) (raw number)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	-2.17	1.00	T-test	0.167
Variance	7.77	18.80	F-test	0.354
SD	2.79	4.34		

6.3.6.4.3 Performance 1 – 4 & 3 – 4 (inter-visit change)

Examining the change in accuracy of performance between P1 and P4, and P3 and P4, examines the change in efficiency between visit one and two respectively. This examines the effect of a hiatus between task performance and feedback.

No significant difference between Groups A and B were identified in relation to inter-visit change in accuracy of performance, as measured using BPC, nor in the consistency of this change (Tables 40 and 41), although the inter-group difference when comparing P1 and P4 performance did approach significance (Table 40, p=0.092). However, again, an interesting difference in trends can be observed. The average gradient of change between P1 and P4 and P3 and P4 in Group A (ITF) was positive (+1.33 and +3.17 respectively). This illustrates that, on average, participants in Group A were less accurate at the start of visit two that they had been at either the beginning or end of visit one. This is seen in Figure 47. However, the average gradient of change between P1 and P4 and P3 and P4 in Group B (DF) was negative (-0.43 and -0.83 respectively), illustrating an improvement in efficacy within Group B when comparing performance at the beginning or end of study visit one with the beginning of visit two.

Table 40: Analysis of BPC change in performance (P1 – P4) (raw number)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	+1.33	-0.43	T-test	0.092
Variance	3.07	2.34	F-test	0.775
SD	1.75	1.53		

Table 41: Analysis of BPC change in performance (P3 – P4) (raw number)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	3.17	-0.83	T-test	0.271
Variance	47.77	22.17	F-test	0.419
SD	6.91	4.71		

6.3.6.5 Task one analysis: Performance four

As interesting trends involving inter-visit differences were observed during the previously detailed quantitative analysis, analysis of performance four (start of visit two) was investigated. The results are summarised in Table 42.

The mean TTC, CIPL and BPC for P4 were all lower and, therefore, more efficient/accurate, in Group B (DF) compared to Group A (ITF). The inter-group difference in performance approached statistical significance in all three measures of performance (Table 42), though the markers of efficiency (TTC and CIPL) were closest ($p=0.096$ and 0.085 respectively).

However, when intra-group variation was compared, statistically significant differences were seen ($p<0.05$) in all three performance measures. These results strongly suggest that participants in Group B (DF) were more consistent (the performances were more clustered) at the beginning of study visit two than their Group A (ITF) counterparts.

Table 42: Task one performance four analysis

		Group A (ITF)	Group B (DF)	P value	
Time taken to Complete (TTC) (seconds)	Mean Variance SD	270.67 41168.67 202.90	100.20 356.99 18.89	T-test F-test	0.096 **<0.001**
Combined instrument path length (CIPL) (metres)	Mean Variance SD	6.77 23.86 4.88	2.41 0.88 0.94	T-test F-test	0.085 **0.002**
Number of badly Placed clips (BPC) (raw number)	Mean Variance SD	5.67 35.47 5.96	2.00 1.20 1.10	T-test F-test	0.198 **0.002**

6.3.7 Task two analysis (TTC)

Limited quantitative analysis of performance in task two was performed via exploration of TTC. Task two was performed twice, once at the end of each study visit one week apart. Figure 44 (section 6.3), illustrates the relative timing of each task two performance. This analysis is summarised in Tables 43 and 44. The pattern of change in mean efficiency measured via TTC is illustrated in Figure 48.

Analysis of task two TTC shows that on average Group B (DF) participants completed task two more quickly than those in Group A (ITF). However, the difference as seen in Figure 48 was not statistically different for either P1 or P2 (Table 43, p=0.495 and 0.963 respectively). What did approach statistical difference, however, was the reduced variation seen in P2 TTC in Group B participants compared to Group A (p=0.100). Variation in Group B performance was five times less than that observed in Group A (Table 49), conferring greater consistency in that group. This pattern is again observed when the variation in the gradient of change between P1 and P2 is examined (Table 44).

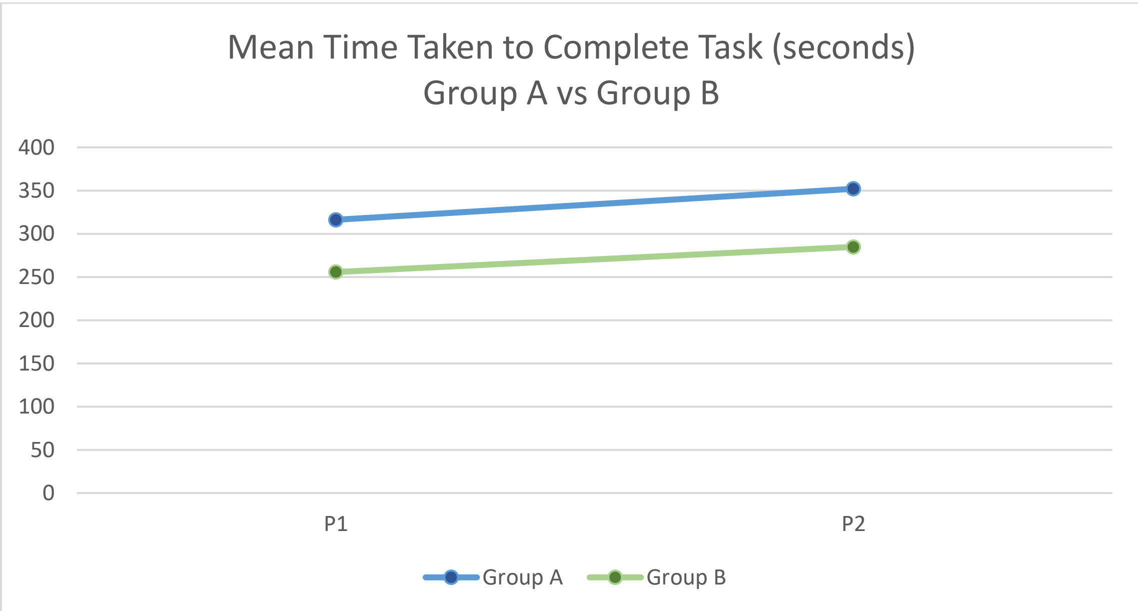
Table 43: Task 2 analysis: TTC

		Group A (ITF)	Group B (DF)		P value
P1 Time taken to Complete (TTC) (seconds)	Mean Variance SD	316.17 21394.17 146.27	255.67 22338.27 149.46	T-test F-test	0.495 0.963
P2 Time taken to Complete (TTC) (seconds)	Mean Variance SD	352.04 57489.11 239.77	296.60 11405.44 106.80	T-test F-test	0.621 0.100

Table 44: Task 2: Analysis of TTC change in performance (P1 – P2)

	Group A (ITF)	Group B (DF)		P value
Mean gradient Variance SD	35.88 21940.99 148.12	40.93 4615.47 67.94	T-test F-test	0.942 0.112

Figure 48: Task 2 mean TTC for P1 – P2, Group A vs B, in seconds



6.3.8 Post-study participant confidence & feedback satisfaction analysis

The post-study participant confidence in self in relation to performing laparoscopic tasks and in the tutor (researcher) in providing useful feedback was measured at the end of this study. As with the pre-study questionnaire, participants were asked to score each of these factors between 1 (low) and 5 (high).

Potential differences between Groups A (ITF) and B (DF) were investigated via an independent Student t-test. The post-study participant confidence in self was not significantly different between groups (Table 45, p=0.662). Analysis beyond calculation of mean was not possible in relation to post-study confidence in tutor and feedback satisfaction as all participants in Group B rated both items maximally. However, both groups appear to have high levels of confidence and satisfaction in the feedback (Table 45). In conclusion, it would appear there was no inter-group differences in relation to these factors with no resulting confounding factor.

Table 45: Post-study confidence and satisfaction analysis of participants

		Group A (ITF)	Group B (DF)		P value
Post-study self-confidence (range 0 – 5)	Mean Variance SD	3.14 0.48 0.69	3.33 0.67 0.82	T-test F-test	0.662 0.687
Post-study tutor-confidence (range 0 – 5)	Mean Variance SD	4.4286 0.29 0.53	5.00 0 0	T-test F-test	N/A N/A
Post-study feedback satisfaction (range 0 – 5)	Mean Variance SD	4.86 0.14 0.38	5 0 0	T-test F-test	N/A N/A

6.4 DISCUSSION

6.4.1 Summary

- Different models of feedback can be orchestrated and creating that difference was supported by use of the custom designed feedback forms.
- The information transfer feedback was highly structured and tutor driven, characterised by a flow of information from tutor to participant. All of the subtasks were discussed sequentially according to the feedback form.
- The dialogic feedback was exploratory in nature; discussion was tutor-facilitated and content was drawn from the learner. Not all tutor observations were discussed.
- Process goals that encourage task-specific self-regulation can be integrated into dialogic feedback sessions by incorporating them into feedback sessions.
- The quality, content and volume of feedback from opposing models was successfully controlled.
- Participants be able to engage with dialogic feedback and, importantly, the less structured dialogic feedback appears no less effective than the highly-structured information transfer feedback

  The quantitative results of this pilot illustrate interesting trends in performance, which suggests a potential difference in the relationship between feedback type and psychomotor task performance

6.4.2 Study design

6.4.2.1 Participants, recruitment and randomisation

The 20 students recruited to this pilot study represents roughly one eighth of Year 3 medical students at Dundee Medical School. Therefore, based upon this response rate, it would appear that participation in the study is relatively attractive to students and email would appear to be an acceptable and successful way to recruit participants. The number recruited was adequate for a pilot study but a larger study would require larger numbers. Therefore, for the final study, I recruited from two year groups and achieved greater flexibility in relation to the timing of participation by organising recruited participants into cohorts, with staggered start weeks.

6.4.2.2 Study protocol

The study protocol was kept close to hand when carrying out study visits and it was strictly adhered to. Keeping a study folder with blank forms and completed forms filed appropriately by section made pre- and post-visit checks easy, which helped ensure data collection was complete. This promoted the integrity and accuracy of the pilot study.

The two pilot studies provided invaluable experience and information pertinent to the final study design. The final study will be longer with an increased number of study visits, providing more opportunities to observe and explore the interesting intra- and inter-visit effects of feedback seen in this pilot study.

The experience of this pilot study and was not only useful in experience and preliminary exploration of the relationship of feedback and its quantitative effect on performance but was also useful in highlighting the limitations of a purely quantitative study design. A quantitative study might be able to explore what the relationship is between type of feedback and psychomotor task performance but it cannot explain why the relationship is thus. For this reason, I decided to add a second research question to the study, specifically exploring how the type of feedback effects participant experience, and why different models of feedback may interact differently with psychomotor skill development and retention. Resultantly, the final study was of a mixed methods design, with a quantitative element concerned with measurement of task performance, and a qualitative element, for exploration of participant understanding and experience.

6.4.2.3 Orientation

It was observed that participants were confused by the Likert scales on the pre-study data collection form. It was decided that future versions of the form should include a clearer explanation of the scale (one conferring the lowest confidence/experience, five conferring the highest confidence/experience). It was also helpful to ask participants to breakdown their previous laparoscopic experience by type of instruments used and duration. Therefore, this section of the pre-study data collection form was amended to simplify completion.

This pilot study highlighted potential conflict with collecting pre-study confidence in tutor (researcher) as part of the same data collection sheet which contained the participant's study ID. If this information is attributable to the individual participant, they may feel pressured into a higher rating than if this were anonymous. Future versions of the post-study data collection form included a separate sheet for recording of this data, with only the study group (A or B) indicated. This information was stored separately to participant-identifiable data and allowed analysis of group averages.

6.4.2.4 Task instructions

The task instruction booklet was successful in giving participants adequate task instructions such that all participants were able to complete the first task performance, in some fashion, without tutor intervention. Variation in the degree to which these instructions were read, however, led to variation in the amount of correction required during the practice session. Therefore, use of an instructional video was considered as a method of controlling the time spent on task directions.

6.4.2.5 Task practice

The length and content of the task practice sessions in visit one and two were well controlled and analysis showed no significant difference in the time spent practicing between Group A (ITF) and B (DF). The controlled task practice sessions were successful in checking participant understanding and encouraging a uniform baseline for participant understanding at the start of the study.

6.4.3 The psychomotor task

6.4.3.1 Task selection and settings

The success of using a simulated laparoscopic task, and specifically the vessel ligation task, observed in the first pilot was repeated in the second pilot study. All participants were able to complete the task and the automatically collected quantitative measures of efficiency and accuracy illustrated similarities and differences between intra-and inter-participant performances. No ceiling effect was observed. The same was true in relation to the second task, vessel sectioning.

6.4.3.2 Quantitative measures

The quantitative measures of task performance selected for inclusion in this pilot study were accurately collected and there was no missing data. Analysis of this data proved interesting and supported their inclusion in future development of the study. The quantitative results in this study provided essential in ensuring a statistically robust design for the final study.

6.4.4 The feedback intervention

In order that the study may be valid, it is vital that there are substantial differences between the models of feedback employed in the two different arms of this study. In this pilot study, the two different types of feedback were consistent with their characteristics as set out in Box 4 (Section 3.6).

The information transfer feedback sessions employed with Group A were highly structured and tutor driven. The tutor worked sequentially through each of the subtasks and gave detailed, specific feedback in relation to each one. A grading based on outcome was awarded for each subtask; good technique was reinforced and detailed directions for how to improve technique were offered for all subtasks and the form was completed in full. Each session ended with a global performance rating. In these sessions, participants were able to clarify points made and offer points to aid tutor understanding but the flow of information was predominantly from tutor to participant.

The dialogic feedback sessions by contrast were far less structured and, in fact, the participants were given opportunity to control the focus of discussion as they were asked “How did that go?” at the start of each session. In these sessions, the participant was encouraged to speak and the tutor facilitated better understanding rather than just telling. The discussion was exploratory in nature, with the participants asked, “what if?” questions when prompts were required for considerations to change in technique. Discussion focused on the understanding of the effect of techniques employed and not on the outcome techniques employed, their success or otherwise. Not all of the tutor observations were always discussed. Asking the Group B participants to formulate two process goals encouraged self-regulation, although it cannot be extrapolated from this that they did self-regulate during procedures. ‘Awareness of self-regulation’ was one of the specific components of the feedback form, which encouraged specific discussion of this skill during feedback sessions.

In this pilot study, efforts to control the length of the feedback sessions between the two study arms were successful. This removed this factor as a potential confounding factor and promotes quality in the study. Similarly, the quality and content of the feedback was controlled between the two study arms by use of pre-entered comments in relation to each subtask or component, which were identical on both feedback forms (Appendices N and O). The use of pre-entered notes on common observations was a useful shortcut for note taking.

In the final study, audio recording of the feedback sessions would allow independent verification of the difference in format and controlled quality and content of the information transfer and dialogic feedback models and sessions respectively.

6.4.5 Results

The quantitative measures of task performance selected for inclusion in this pilot study were accurately collected and there was no missing data. Analysis of this data proved interesting and supported their inclusion in future development of the study.

The quantitative measures of efficiency - Time Taken to Complete the task (TTC) and Combined Instrument Path Length (CIPL) - showed that efficiency improved over the length of the study in both groups. However, more interestingly, the pattern of this change was similar between both groups in both measures:

- Both groups improved during visit one
- Group A performance deteriorated whilst Group B performance improved at the start of the second study visit

- Group A performance improved whilst Group B performance deteriorated during the second study visit

Therefore, analysis of these performance markers is of interest and, in particular, the relationship between type of feedback and intra-visit and inter-visit efficiency is worthy of further exploration.

The quantitative measure of accuracy – number of Badly Placed Clips (BPC) – showed a different trend over the length of the study for Group A and Group B; Group A became less accurate over the course of the study, with a higher average number of BPC in Performance 5 than Performance 1, but in contrast, Group B became more accurate. However, the same pattern in relation to study visit two was similar to that observed in relation to efficiency:

- Group A performance deteriorated whilst Group B performance improved at the start of the second study visit
- Group A performance improved whilst Group B performance deteriorated during the second study visit

Furthermore, whilst the purpose of this pilot study was to explore feasibility and suitability of the task, rather than to investigate the relationship between feedback and performance, a statistically significant inter-group difference in relation to consistency of performance (all performance measures) at the start of the second study visit (P4) was illustrated. This was also seen in relation to the second task TTC. Given that statistical analysis is limited by the naturally high variation in quantitative measures of performance, this is a noteworthy finding.

CHAPTER 7: FINAL STUDY METHODS

7.1 RESEARCH QUESTIONS

As per section 3.8, I proposed two research questions for the investigation of dialogic feedback with encouraged self-regulation:

1. What is the effect of an integrated model of dialogic feedback with encouraged self-regulation versus an information transfer model of feedback on psychomotor task performance and longevity of skill retention?

2. What is the effect of the integrated model of dialogic feedback model on learner's experience and understanding of feedback?

7.2 ETHICAL APPROVAL

Ethical approval for this study was sought prior to commencement of the pilot studies. The University of Dundee Research Ethics Committee (Application 14134) granted permission on 29th October 2014. (Appendix A)

7.3 STUDY DESIGN

7.3.1 Sample size calculations

The second pilot study data was valuable in informing sample size calculations for the final study. These calculations were based on analysis of the data from task performance 4 (at the start of the second study visit) as an inter-group difference in variation in performance was noted at this point in the study in all three measures of performance (time taken to complete, economy of movement and errors made in relation to number of badly placed clips). Adequately powering the study in relation to as many quantitative measures as possible resulted in a more statistically robust study.

A series of separate sample size calculations were performed based upon the separate performance markers from the second pilot study results (TTC, CIPL and BPC) using the 'mean difference sample size' calculator available at the Open Source Epidemiologic Statistics for Public Health website (http://www.openepi.com). Significance was set at 95% and power to 80%. Calculations were performed in two different ways in order to explore the balance of feasibility and rigour:

1) The sample size calculation was based upon purely the pilot experimental data only, using the inter-group difference and standard deviation.
2) The sample size calculation was based upon the standard deviation from experimental data plus a pre-set minimally significant inter-group difference.

This series of calculations are attached as Appendix P. Based upon these findings, the final study sample size was set as 30 participants in each group, 60 in total. This sample size will be sufficient to detect significant inter-group differences in task performance across measures of both efficiency (time taken to complete and economy of movement) and accuracy (number of badly placed clips).

7.3.2 Participants & recruitment

Fourth and fifth year undergraduate medical students at the University of Dundee were recruited to this study via six invitational emails to their year groups. In return for participation in the study, the students were offered a certificate of participation for their undergraduate portfolio and a Dundee Medical School Green card; a record submitted to the medical school office to reflect a student's contribution to extra-curricular activities. This study was conducted over seven months between 28th November 2014 and 29th June 2015.

Recruiting medical students to a medical education study ensured the validity of the participant group. Medical students were the most commonly recruited participants to the experimental feedback studies identified in the structured literature review (Chapter 2, Section 2.3.3). The laparoscopic experience of senior medical students is relatively homogenous, minimising pre-study experience as a confounding factor. There is also a large number to recruit from (approximately 300 in both years), making recruiting large numbers for an adequately statistically powered study more feasible. Seventy-two medical students were recruited in total, with 4 participants ultimately failing to start the study and seven failing to attend one of the study visits and subsequently being removed from the study. Therefore, a total of sixty-one participants completed the study.

7.3.3 Participant randomisation

Participants were organised by recruitment group (dependent on timing of volunteering) and randomised to Group A (information transfer feedback) or Group B (dialogic feedback). Using the 'random number' function in Microsoft Excel, each participant was allocated a random number, ranked in numerical order by this number and alternately allocated a study number in Group A or B. This process maintained relatively equal numbers in each group throughout the study. The study numbers of the eleven of the seventy-two recruited participants that either did not start or dropped out of the study were reallocated. The file depicting this randomisation process can be found in the accompanying data disk ('Full study Participant randomisation').

The final participant group included thirty-one randomised to Group A and thirty randomised to Group B.

7.3.4 Study setting

The final study was based at the Tayside Orthopaedic & Rehabilitation Technology (TORT) centre at Ninewells Hospital & Medical School. For the purpose of the final study, one of the Cuschieri Skills Centre LapSim® models was moved to an office within the TORT centre.

7.3.5 Study protocol

Figures 49-52 refer to the study protocols employed in visit one, two, three and four respectively. Each study visit was structured as per the 'session components' described.

Figure 49: Visit one protocol (week one)

Session component	**Participant group**	
	Group A; information transfer feedback (n = 31)	*Group B; dialogic feedback with self-regulation (n = 30)*
Orientation	Study information sheet issued; Consent form issued / signed; Pre-study data collection sheet completed	
Explanation of task	Participant view Task 1 instruction video. Opportunity for participant questions.	
Task practice	5 minutes of supervised and structured Task 1 practice Tutor giving instruction on how to use LapSim controls, aims of the task, clip applying technique but NOT tactics or technique correction	
Task performance	**Task 1 performance 1 (T1/P1)**	
Feedback session 1	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Task performance	**Task 1 performance 2 (T1/P2)**	
Feedback session 2	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Task performance	**Task 1 performance 3 (T1/P3)**	
Feedback session 3	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Debrief	Thank for time; arrange follow up in approximately 7 days	

Figure 50: Visit two protocol (Week two)

Session component	**Participant group** *Group A; information transfer feedback*	 *Group B; dialogic feedback with self-regulation*
Task practice	5 minutes of supervised unstructured Task 1 practice No tutor instructions	
Task performance	**Task 1 performance 4 (T1/P4)**	
Feedback session 4	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Task performance	**Task 1 performance 5 (T1/P5)**	
Feedback session 5	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Debrief	Thank for time; arrange follow up in approximately 7 days	

Figure 51: Visit three protocol (Week 3)

Session component	**Participant group** *Group A; information transfer feedback*	 *Group B; dialogic feedback with self-regulation*
Task practice	5 minutes of supervised unstructured Task 1 practice No tutor instructions	
Task performance	**Task 1 performance 6 (T1/P6)**	
Feedback session 6	- Information transfer feedback with global rating scale	- Dialogic feedback with self-regulation skill development - Set 2 goals for next task completion
Task performance	**Task 1 performance 7 (T1/P7)**	
Microanalysis interview	Tutor/participant review of recorded T1/P7 performance with microanalysis questionnaire	
Explanation of task	Participant view Task 2 instruction video. Opportunity for participant questions.	
Task performance	**Task 2 performance 1 (T2/P1)**	
Debrief	Thank for time; arrange follow up in approximately 4 weeks	

Figure 52: Visit four protocol (Week seven)

Session component	Participant group *Group A; information transfer feedback*	 *Group B; dialogic feedback with self-regulation*
Task performance	**Task 1 performance 8 (T1/P8)**	
Task performance	**Task 2 performance 2 (T2/P2)**	
Structured interview	Semi-structured interview exploring learner understanding and experience of feedback	
Debrief	Thank the participant for their time Pre-study data collection sheet completed	

In this study, the participants assume the role of the learner and the researcher the role of the tutor. All study visits and feedback sessions were conducted by myself, the lead researcher.

7.3.6 Study visits

In this final study, the number of study visits was increased to four as this would increase the opportunity to observe the intra- and inter-visit effect, i.e. the immediate and medium-term effect, of feedback on psychomotor task performance. As per the second pilot study, the inter-visit gap between the first three study visits was seven days. With little consensus in the medical education literature, this inter-visit gap was based on the clinical experience of operating lists often being weekly occurrences.

The effect of a time elapse between the feedback intervention and task performance is of interest in this study and is specifically relevant to my first research question. Three of the studies identified via the structured experimental studies feedback literature review (Chapter 2; Xeroulis et al, 2007; Porte et al, 2007; Kruglikova et al, 2010) featured a delayed skill retention test. All of these studies chose a period of 4-6 weeks between post-intervention and delayed testing. Therefore, to aid comparison of results between this study and existing literature, a delayed skill retention test was incorporated into the study design. This fourth and final study visit would take place as close to 28 days after the third study visit as possible, within the limitations of participant availability.

7.3.7 Orientation

At the start of visit one, participants were provided with a study information sheet (Appendix Q) and informed consent (Appendix R) was obtained. Participant information was also collected (Appendix S). This included age, sex, dominant hand, self-rated confidence level, self-rated experience level, description of experience and participant confidence in tutor (researcher).

7.3.8 Participant task instructions

Participant instructions were given via two task instruction videos. The task one video was viewed during visit one, immediately prior to task practice. The task two video was viewed during visit three, immediately prior to the first performance of task two. These videos are available to view on the data disk ('Task one instructional video', 'Task two instructional video'). This method of imparting task instruction allowed the instructions to be standardised between participants, promoting quality of the study. Video instructions allowed greater detail to be incorporated into the instructions without them becoming difficult to digest, as might have been the case with more detailed written instructions. Also, the pace of the delivery of these instructions was controlled; participants could not 'skip through' sections as one might when reading.

7.3.9 Task practice

As with the second pilot study, participants were given an opportunity to practice task one prior to the first task performance in each study visit (T1/P1, T1/P4, T1/P6), with the exception of visit four. Practice time was limited to approximately five minutes. The aim of this brief, time-controlled practice was to standardise participant understanding at the start of the study and each subsequent visit and, therefore, remove the effect of variation in understanding as a reason for variation in performance during the study.

During task one practice, participants were supervised and asked to complete five actions: changing instrument type, grasping and moving the vessel, attempted application of a clip, rupture of the vessel via excessive tension, and use of the suction device. This promoted an even baseline of participant competence prior to data collection and, thus, promoted the quality of the study.

As the purpose of task two was to assess the participant's ability to apply skills acquired during task one performance and feedback to another task, the very first performance of that task was of interest. No opportunity was given for task practice and correction of understanding and technique.

The purpose of the fourth study visit was to assess participant skill retention after a four-week break from study visits. To increase the sensitivity of this assessment, there was no practice scheduled in relation to task one or two and participants proceeded immediately to task performances (T1/P8, T2/P2).

7.3.10 Quantitative design elements

7.3.10.1 Task one performance and repetition

Task one (vessel ligation and division) was performed three times during visit one (T1/P1, T1/P2 and T1/P3), twice during visits two (T1/P4 and T1/P5) and three (T1/P6 and T1/P7) and once during visit four (T1/P8). The increased number of study visits and task one repetitions allowed for more detailed examination of the changes and trends in intra- and inter-visit performance over the course of the study. Comparison of performances during the same study visits (T1/P1, T1/P2 and T1/P3, T1/P4 and T1/P5, and T1/P6 and T1/P7) quantified the immediate effect of feedback (the intra-visit feedback effect). Comparison of performances from different study visits (T1/P3 and T1/P4, T1/P5 and T1/P6 and T1/P7 and T1/P8) illustrated the effect of feedback in the context of a hiatus on participant performance (the inter-visit feedback effect).

7.3.10.2 Task two performance and repetition

Task two (vessel sectioning) was performed during study visits three (T2/P1) and four (T2/P2). Analysis of T2/P1 reflects the ability of participants to transfer simulated laparoscopic motor and metacognitive skills from one task to another. Comparison of T2/P1 and T2/P2 investigated effect feedback in the context of a hiatus on participant performance (the inter-visit feedback effect).

7.3.11 Qualitative design element: Semi-structured interview

The second qualitative element of this study comprises a semi-structured participant interview in which participants' understanding and experience of feedback is explored. This interview was included in study visit four (week seven). It was deliberately conducted four weeks after the last feedback session (visit three, week three) as I wanted to capture participant ideas and opinions formed after an opportunity to reflect on their experiences gained during the study, rather than the immediate reactions that would be more likely represented in an interview conducted earlier in the study.

The interview comprised of nine core questions (Box 11), with additional explorative discussion of the ideas expressed. The interviews were audio recorded and transcribed prior to thematic analysis.

Box 11: Semi-structured participant interview questions (visit four)

1 Having been involved in the study, what do you think the purpose of feedback is?

2 How did the feedback process in this project help you learn?

3 Having participated in the study, have you taken anything away that might help you learn in future?

4 How important is setting aims or goals when performing a task?

5 What attributes does a useful aim or goal have? Give me an example.

6 During the study, how did you assess progress? (Group A: Did you use the global rating scale?)

7 Who or what was the main source of feedback during the study? Were there any other sources?

8 Did you enjoy the feedback sessions? And why?

9 Was there anything you found difficult or unenjoyable about the feedback?

(Group A: I'm aware I did a lot of the talking; would it have been helpful for you to speak more?)

7.3.12 Study debrief

At the end of visit four, participants were asked to complete the post study data collection form (Appendix T), in which they rated their post-study self confidence in performing laparoscopic tasks and participant confidence in the tutor (researcher) in providing useful feedback. Analysis of this data investigated the relationship between feedback type and participant self-confidence and tutor confidence. Participants were asked to rate their overall satisfaction with the feedback provided as inter-group variation in these ratings may have resulted in a potential source of bias. Participants were thanked for participating in the study and arrangements were made for collection of participation certificates.

7.4 THE PSYCHOMOTOR TASK

7.4.1 Task one: Clip ligation and division

The ‘clip applying’ task on the LapSim® model was employed as the first task in this study. Experience of using this task in the two pilot studies provided data on which to base task settings, added to my experience of teaching it and to development of the feedback intervention. Details of the task purpose and technique can be reviewed in chapter 5, section 5.2.2.

7.4.1.1 Task one settings

The task-specific simulator settings in relation to task one are illustrated in Box 12. These settings were unchanged from the pilot studies.

Box 12: Task one LapSim® ‘Clip application’ task settings	
Camera angle	-20 degrees (vessel slightly oblique to viewer, left side closer)
Clip target area size (mm)	4
Cutting target area size (mm)	4
Stretch sensitivity	Low
Spontaneous bleeding	Off
Bleeding flow rate (L/min)	1

7.4.1.2 Task one measures of performance

The quantitative measures of task performance collected and analysed are detailed in Box 13. These measures assess participant efficiency and accuracy.

Box 13: Task one LapSim® 'Clip application' measures of performance

Measure	Abbreviation	Units	Efficiency or accuracy
Total time to complete task	*TTC*	*Seconds*	*Efficiency*
Combined (right and left) instrument path length	*CIPL*	*Metres*	*Efficiency*
Number of incomplete (missed) target areas	*ITA*	*Raw number, range 0-3*	*Accuracy*
Number of badly placed clips	*BPC*	*Raw number*	*Accuracy*

Each quantitative measure of performance is crude when separated into individual values. It is almost artificial to discuss performance in terms of these separate measures but no single objective measure of performance exists. Creation of a global rating would be complex, impossible to validate and at risk of subjectivity. Therefore, analysis of multiple objective quantitative measures, and assimilation of this information in interpretation of the whole picture, is of value.

7.4.2 Task two: Vessel sectioning

The 'vessel sectioning' task on the LapSim® model was employed as the second, 'cross-over' task in this study. Experience of using this task in the second pilot study provided data on which to base task settings. Details of the task purpose and technique can be reviewed in chapter 6, section 6.2.2.2.

7.4.2.1 Task two settings

The task-specific simulator settings in relation to task two are illustrated in Box 14. These settings were unchanged from the second pilot study.

Box 14: Task two LapSim® 'Clip application' task settings

Camera angle	-10 degrees (vessel slightly oblique to viewer, left side closer)
Camera type	Non-moving
Clip target area size	Medium
Cutting target area size	Medium
Stretch sensitivity	Low

7.4.2.2 Task two measures of performance

The quantitative measures of task performance collected and analysed are detailed in Box 15. These measures assess participant efficiency and accuracy.

Box 15: Task two LapSim® 'Clip application' measures of performance

Measure	**Abbreviation**	**Units**	**Efficiency or accuracy**
Total time to complete task	*TTC*	*Seconds*	*Efficiency*
Combined (right and left) instrument path length	*CIPL*	*Metres*	*Efficiency*
Frequency of abdominal wall damage	*AWD*	*Raw number*	*Accuracy*
Maximum depth of abdominal wall damage	*MDD*	*Millimetres*	*Accuracy*

7.5 THE FEEDBACK INTERVENTION

Feedback forms for each of the feedback group were based upon the characteristics of information transfer and dialogic feedback as described in chapter 3 (Section 3.7, box 9) and were revised based upon the experience gained during the pilot studies.

Attempts were made to control the time spent on each feedback episode to between 5 and 10 minutes. The time spent on each episode was recorded for analysis.

7.5.1 Information transfer feedback

The revised information transfer feedback form is shown below in Figure 53 and is attached as Appendix U. This form details global and task specific elements, facilitating organised and comprehensive note taking during tutor observation of performance and structure and detail of subsequent feedback.

During Group A feedback sessions, the information transmission feedback form was discussed and the task component ratings were completed in full. The researcher/tutor worked through each of the global and task-specific components in order and discussed their observational comments. The participants were awarded a rating for each component: not seen (N), development needed (D) or satisfactory (S). The feedback session concluded with the tutor providing an overall rating of the performance via the global rating (0-4). These feedback sessions were highly structured and the discussion was tutor-led.

Appendix JJ

Date:

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education
Pilot study: Group A
Feedback Type A checklist

Participant number: Performance #: Time for FB:

	Component	**Rating**			**Comments**
G1	Follows and agreed, logical sequence or protocol for the procedure	N	D	S	> At start > Adaptation > At end – bleeding / clip removal > Problem solving
G2	Consistently handles tissue well with minimal damage	N	D	S	> Arm position > Danger to vessel when 2 instruments being used > Inadvertent movements > Ruptured vessel
G3	Controls bleeding promptly	N	D	S	> Is 1st thought suction > Clipping held vessel before transferring to 2nd bleeder > Use of grasper for temporary arrest > Recognising bleeding vessel
G4	Uses instruments appropriately and safely	N	D	S	> Instrument selection (Early) > Dexterity > Instruments always in view > Orientation > Co-ordination of instruments
G5	Proceeds at an appropriate pace with economy of movement	N	D	S	> Bleeding - complete blood removal before changing task > Check clip good before releasing vessel > Multitasking with 2 instruments > Reducing instrument travel > 2nd hand neglect
G6	Deals calmly and effectively with unexpected events and complications	N	D	S	> Emotional reaction > Tactical reaction
T1	Grasps vessel carefully with grasper to allow clip application	N	D	S	> Position of grasper on vessel > Positioning of vessel > Holding vessel steady
T2	Applies clips correctly to either side of vessel	N	D	S	> Position of clip - Within target areas - To stop bleeding - Taught not flaccid vessel end > Optimising clip technique - Adequate tension - Vessel properly in jaws - Seeing both jaws of clipper - Orientation of clipper
T3	Cuts vessel safely with scissors	N	D	S	
T4	Retrieves dropped clips	N	D	S	> Use of 2 graspers > Use of closest grasper > Dexterity to reduce accidental clip closure

Level of performed task		
0	Unable to complete task	
1	Able to complete task with significant difficulty in several components	
2	Able to complete task with mild-moderate difficulty in several components	
3	Able to perform all elements of task with mild difficulty in only some components	
4	Able to perform all elements of task, exhibiting high skill level	

Figure 53: Information transfer feedback form

7.5.2 Dialogic feedback

The revised dialogic feedback form is shown below in Figure 54 and is attached as Appendix V. This form retained the global and task-specific elements from the original form as these were successful in organising comprehensive and detailed feedback for discussion during the dialogic feedback sessions. Observed behaviours was separated into ‘systematic/tactical awareness’ and ‘technical skill’ domains to help emphasise awareness of these spheres, rather than focus purely on task components. Discussion and recording of two process goals was incorporated into the form to promote the feed-forward self-regulatory element of the feedback sessions. These process goals related to two behaviours that they would like to use in their next task performance. Participants were reminded of their process goals prior to each task repetition and reflection on these goals formed part of the subsequent feedback session. The purpose of the dialogic feedback form was to promote discussion of process rather than record outcomes and judgement.

Group B feedback sessions were explorative and discussion focused on participant understanding of process rather than assessment of outcome.

Appendix KK

Date:

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Feedback Type B Form / notes

Participant number: Performance #: Time taken for FB:

	Component	**Systematic / tactical awareness**	**Technical skills**
G1	Logical sequencing	> At start > Adaptation > At end – bleeding / clip removal > Problem solving	
G2	Tissue handling	> Arm position > Danger to vessel when 2 instruments being used > Inadvertent movements > Ruptured vessel	
G3	Use of instruments	> Instrument selection (Early)	> Dexterity > Instruments always in view > Orientation > Co-ordination of instruments
G4	Pace with economy of movement	> Bleeding - complete blood removal before changing task > Check clip good before releasing vessel	> Multitasking with 2 instruments > Reducing instrument travel > 2nd hand neglect
G5	Deals calmly and effectively with unexpected events and complications	> Emotional reaction > Tactical reaction	
G6	Awareness of self-regulation	> Were you monitoring your performance? > How were you monitoring? What were you checking for? > Was there evidence of your goals/aims in this performance? > Evidence of checking work	
T1	Use of grasper to allow clip application	> Position of grasper on vessel > Positioning of vessel	> Holding vessel steady, reducing accidental movement
T2	Applies clips correctly	> Position of clip - Within target areas - To stop bleeding - Taught not flaccid vessel end	> Optimising clip technique - Adequate tension - Vessel properly in jaws - Seeing both jaws of clipper - Orientation of clipper > "Sneaking up"
T3	Control of bleeding	> Is 1st thought suction > Clipping held vessel before transferring to 2nd bleeder > Use of grasper for temporary arrest	> Recognising bleeding vessel
T4	Minimises dropped clips / Optimises method of retrieval	> Use of 2 graspers > Use of closest grasper > Awareness of number dropped	> Dexterity to reduce accidental clip closure

GOALS/AIMS: 1.

2.

Figure 54: Dialogic feedback form

7.6 METHODS OF STATISTICAL ANALYSIS

7.6.1 Adherence to study protocol

An inter-group comparison was performed in relation to inter-visit intervals, practice time and the volume of feedback provided or facilitated. These comparisons were performed using the student t-test (independent samples, two tailed probability with assumption of unequal variances). For the purposes of statistical analysis, significance was set at 0.05.

7.6.2 Participant demographics, experience and confidence

An inter-group comparison was performed in relation to participant demographics, pre-study experience, and pre- and post-study confidence. Factors measured via continuous data (participate age, pre-study experience, and pre- and post-study confidence) were analysed using the student t-test (independent samples, two tailed probability with assumption of unequal variances). Factors measured using categorical data (participant sex and hand dominance) were analysed using Fishers exact test. For the purposes of statistical analysis, significance was set at 0.05.

7.6.3 Definition & identification of outliers

The purpose of the definition and identification of outliers in performance is to improve the accuracy of the quantitative analysis of the study data and thus promote accurate conclusions, by removing spurious results that may skew statistical analysis. Two potential methods were considered: identification of outlier quantitative measures; and identification of outlying participants.

In this study, absolute quantitative values, the trend of these values and the spread of these values (the latter relating to intra-group consistency) were of interest. Removing all outlying values would reduce the range of values and blunt inter-group statistical comparisons. For this reason, a decision was taken to identify and exclude only participants whose performances were *consistently* statistically different from the group, removing all results relating to those participants.

This analysis identified two outlying participants in each study group (four in total; Data disk, filename ‘Full study Results’, sheet ‘outlier identification’) and the study data correlating to these participants was excluded from further analysis.

7.6.4 Task performance

During each task one performance (P1-8), four quantitative measures of performance were collected (Box 13). During each task two performance (P1-2), another four quantitative measures of performance were collected (Box 15). Figures 49-52 in section 7.3.5 illustrates the timing of study visits and task performances.

7.6.4.1 Direct comparison of performance

An inter-group comparison of each performance (task one P1-8; task two P1-2) was performed using the Student t-test (independent samples, two tailed probability with assumption of unequal variances) in respect to each of the quantitative measures of performance. For the purposes of statistical analysis, significance was set at 0.05. This method both directly tests the quantitative measure of performance (T-test p value) and the variation in this quantitative measure (T-test p value), the latter being a comparison of consistency of performance within the groups.

7.6.4.2 Comparison of performance trend

The trend of each group’s performance, between specific points in the study or across the whole study duration, is calculated via the summation of the gradients of the slopes between the data points of interest for each participant in each group.

Inter-group comparison of trends in performance were made via the Student t-test (independent samples, two tailed probability with assumption of unequal variances). Again, this method both directly tests the trend in the quantified measure of performance in question (T-test p value) and the variation in this trend (F-test p value), conferring the intra-group consistency of this trend.

When calculating these trends in relation to each measure of task performance (Box 1 and 2), with the exception of ‘number of sections obtained’ (task two), a negative gradient confers an improvement in performance. Conversely, a positive gradient confers a deterioration in performance.

7.6.4.2.1 Task one comparison of intra-visit performance trend

In relation to task one, examining the change in measures of performance between P1 and P3, and P4 and P5, and P6 and P7 examines the change in performance during each study visit (visits one, two and three respectively). These results illustrate the intra-visit, immediate effect of feedback on task performance.

7.6.4.2.2 Comparison of inter-visit performance trend

In relation to task one, examining the change in measures of performance between P3 and P4, P5 and P6, and P7 and P8 examines the change in performance between each study visit (visit one and two, visit two and three, and visit three and four respectively). These results illustrate the inter-visit, delayed effect of feedback on task performance. This inter-visit comparison is also applicable to task two analysis, comparing P1 and P2 performance.

CHAPTER 8: FINAL STUDY QUANTITATIVE RESULTS

The raw data and analysis for this study are available for review in the accompanying ‘Data disk’ (filename ‘Full study Raw data’ and ‘Full study Results’ respectively). For the purposes of clarity, the two study groups will be referred to as Group A (information transfer feedback, ITF) and Group B (dialogic feedback, DF) throughout this results section.

8.1 EXAMINING PROTOCOL, DEMOGRAPHICS, CONFIDENCE AND EXPERIENCE

8.1.1 Analysis of integrity of feedback intervention

As the aim of this study was to investigate the effect of dialogic feedback with encouraged self-regulation of learning on psychomotor task performance, skill retention and the learner experience of feedback, it was vital to the validity and reliability of the results that the feedback interventions, via tutor(researcher):participant feedback sessions, was authentic to the models of feedback proposed. The contrasting characteristics of both the information transfer and dialogic feedback models are described in Box 4, reproduced below:

Box 4: Feedback models and their characteristics

Features of information transfer feedback	Features of dialogic feedback with self-regulation
Tutor driven	Tutor facilitated
Content largely given by tutor	Content largely drawn from learner
Learner role is passive	Learner role is active
Discussion is directive	Discussion is exploratory
Focus is on outcome of actions	Focus is on process
Goal-orientated behaviour not promoted	Process goal-orientated behaviour promoted

A random sample of feedback sessions was generated using the ‘random number’ function in the Microsoft Excel programme (see data disc; file ‘Full study Results’; sheet ‘Feedback session review’). The audio recordings of these feedback sessions were then reviewed against the criteria as defined in Box 4.

The result of this review was that the feedback sessions were found to be congruent with the appropriate feedback model for their relevant study group. The information transfer feedback sessions are highly structured, and tutor dominated and driven. The agenda is fixed, the detailed content is contributed almost entirely by the tutor and focus is placed upon the outcome of observed actions. The dialogic feedback sessions are tutor fascilitated but the learner (participant) role is more active and they contribute significantly to the feedback discussion. There is clear encouragement of process goal use and the exploratory discussion focuses on task understanding.

The audio recordings of these feedback sessions are available via the data disc (see data disc; file 'Full study Results'; sheet 'Feedback session review' for instructions regarding access).

8.1.2 Analysis of inter-visit interval

Due to restrictions of availability, it was inevitable that there would be variation in the interval between study visits for different participants. The inter-visit intervals were investigated via an independent Student t-test (two tailed probability with assumption of unequal variances) and analysis is shown in Table 46. As per the study design (chapter 7, section 7.3.6), the intended inter-visit interval (in days) is also given. There was no scientific or statistically significant difference between the inter-visit intervals between Groups A (ITF) and B (DF). The closeness of the average inter-visit interval and the intended inter-visit interval, for each group at each study stage, indicates close adherence to the study protocol and promotes quality of the study.

Table 46: Inter-visit intervals (in days)

		Group A (ITF)	Group B (DF)	P value
Visit 1 – 2 *Intended interval = 7 days*	Mean Median	7.65 8	6.73 7	0.513
Visit 2 – 3 *Intended interval = 7 days*	Mean Median	7.03 7	7.53 7	0.204
Visit 3 – 4 *Intended interval = 28*	Mean Median	28.46 28	28.03 28	0.813

8.1.3 Analysis of practice time

The time Group A (ITF) and B (DF) participants spent practicing at the start of each study visit was investigated via an independent Student t-test (two tailed probability with assumption of unequal variances) and analysis is shown in Table 47. The data shows that Group B spent less time practicing than Group A in visit one (317 vs 337 seconds respectively, p=0.012). However, the significance of a 20 second inter-group difference is unknown. No significant differences were identified in relation to practice time during visit two and three.

Table 47: Analysis of time spent practicing (in seconds)

		Group A (ITF)	Group B (DF)	P value
Visit 1	Mean (SD)	337 (23.90)	317 (34.66)	**0.012**
Visit 2	Mean (SD)	215 (81.62)	212 (76.74)	0.879
Visit 3	Mean (SD)	195 (69.33)	197 (70.44)	0.891

8.1.4 Analysis of feedback volume

The total amount of feedback given to participants in each group (the sum of the length of feedback sessions 1 - 6) was controlled to limit this as a confounding factor between participants and the study groups and promote the quality of the study design. All feedback sessions were timed.

Potential inter-group differences were investigated via an independent Student t-test (two tailed probability with assumption of unequal variances). There was no significant difference in either the volume of feedback provided to either group or the variation in volume of feedback provided (Table 48). There was reduced variation in feedback session length in Group A, which is consistent with the information transfer feedback being more structured, but the difference was not statistically significant.

Table 48: Analysis of volume of feedback sessions (seconds)

	Group A (ITF)	Group B (DF)			P values
Mean total volume (SD)	2608.68	2686.60		T-test	0.423
Variance	106717	176679		F-test	0.176

8.1.5 Demographics

Inter-group comparisons in relation to hand dominance and gender (via Fisher's exact test) and age (via a two-tailed independent Student t-test) were performed (Table 49). There were no significant inter-group differences in relation to handiness or age. However, there was an inequality with respect to gender (p=0.049), with a greater proportion of male participants randomised to Group A (ITF) compared to Group B (DF) (p=0.49).

Table 49: Participant demographics

	Group A (ITF)	Group B (DF)	P values
Hand dominance			
Right	28	3	0.610
Left	29	1	
Gender			
Female	18	25	**0.049**
Male	13	5	
Age (years)			
Mean (SD)	23.71 (2.19)	23.87 (3.21)	0.830

8.1.6 Pre-study participant confidence

The pre-study participant self-confidence in relation to performing laparoscopic tasks and in the tutor (researcher) in providing useful feedback in relation to laparoscopic tasks was via a Likert scale. Participants were asked to score each of these factors from 1 (conferring the lowest confidence) to 5 (conferring the highest confidence). The results were investigated via an independent Student t-test; no significant differences were found (Table 50). The results illustrate a relatively low self-rating of pre-study confidence (Group A mean = 1.81 and Group

B mean = 1.63) but a high pre-study confidence in tutor feedback (Group A mean = 4.03 and Group B mean = 4.20).

Table 50: Analysis of participant pre-study confidence

		Group A (ITF)	Group B (DF)	P value
Pre-study self-confidence (range 1-5)	Mean (SD)	1.81 (0.75)	1.63 (0.72)	0.361
Pre-study tutor-confidence (range 1-5)	Mean (SD)	4.03 (0.59)	4.20 (0.63)	0.255

8.1.7 Pre-study participant experience

Participants were asked to rate their pre-study experience with laparoscopic tasks and instruments on a Likert scale of 1 (conferring the least experience) to 5 (conferring the most experience). Potential differences between Groups A and B were investigated via an independent Student t-test and no significant difference was found (Table 51). The results illustrate a relatively low self-rating of pre-study experience (Group A mean = 1.83 and Group B mean = 2.03).

Table 51: Analysis of participant pre-study experience rating

		Group A (ITF)	Group B (DF)	P value
Pre-study experience rating (range 1-5)	Mean (SD)	1.83 (0.52)	2.03 (0.76)	0.253

As an additional measure of pre-study experience, participants were asked to record their previous laparoscopic experience. Each episode of experience was recorded by type of instruments used (real instruments in theatre, laparoscopic box trainer or virtual reality laparoscopic simulator) and by length of experience in minutes. The total volume of pre-study experience was calculated as the sum of all episodes of experience. Potential inter-group differences were investigated via an independent Student t-test of the total volume of experience; no significant difference was found (Table 52). Seven participants in each group had had pre-study experience of using a virtual reality laparoscopic simulator.

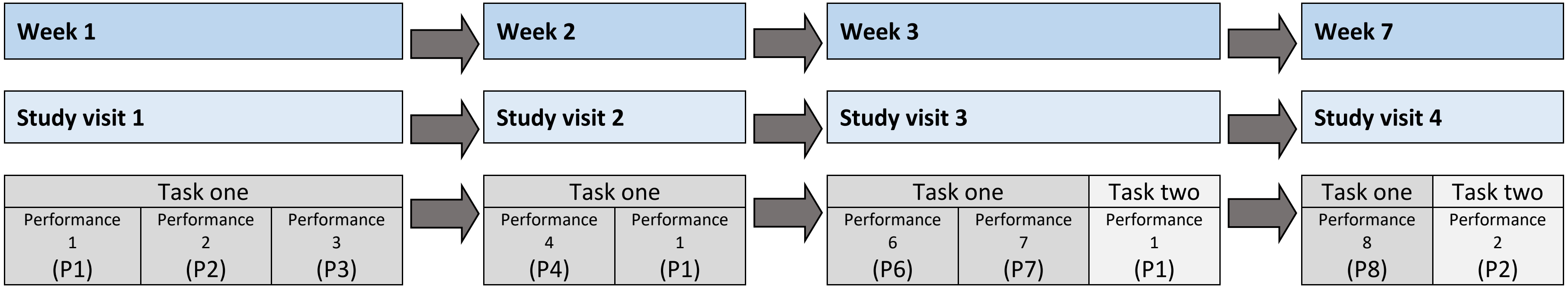
Table 52: Analysis of participant pre-study volume of experience (minutes)

		Group A (ITF)	**Group B (DF)**	**P value**
Pre-study experience volume	Mean (SD)	74.19 (114.93)	97.30 (180.48)	0.555

8.2 TASK ONE PERFORMANCE ANALYSIS (VESSEL LIGATION)

Figure 55 illustrates the timing of study visits and task performances and acts as a reference during interpretation of this results section.

Figure 55: Chart illustrating timing of study visits and task one and task two performances



8.2.1 Intra-group performance improvement (P8 vs P1)

Both study groups demonstrated significant improvement in performance with respect to all four measures of efficiency (time taken to complete, TTC, and combined instrument path length, CIPL) and accuracy (number of badly placed clips, BPC, and incomplete target areas, ITA) over the course of the study (P8 vs P1 performance, Table 53). This suggests that both information-transfer and dialogic models of feedback are associated with improved performance of psychomotor tasks. Both groups also exhibited less intra-group variation during P8 as compared to P1, suggestion an association between both models of feedback and improving consistency of performance.

Table 39: Intra-group performance analysis (P8 vs P1)

		Group A (P1)	Group A (P8)		P values
Time taken to Complete (TTC) (seconds)	Mean Variance SD	394.77 71407.31 267.22	172.12 9150.79 97.55	T-test F-test	**<0.001** **<0.001**
Combined instrument path length (CIPL) (metres)	Mean Variance SD	705.39 1592.71 398.97	282.27 29535.35 171.86	T-test F-test	**<0.001** **<0.001**
Number of Incomplete Target Areas (raw number, range 0-3)	Mean Variance SD	2.26 1.26 1.13	0.42 0.71 0.84	T-test F-test	**<0.001** 0.146
Number of badly Placed clips (BPC) (raw number)	Mean Variance SD	5.323 27.69 5.26	1.73 4.50 2.12	T-test F-test	**<0.001** **<0.001**
		Group B (P1)	**Group B (P8)**		**P values**
Time taken to Complete (TTC) (seconds)	Mean Variance SD	316.43 25387.77 159.34	133.52 4786.18 70.50	T-test F-test	**<0.001** **<0.001**
Combined instrument path length (CIPL) (metres)	Mean Variance SD	621.50 146364.74 382.58	219.61 13347.24 115.53	T-test F-test	**<0.001** **<0.001**
Number of Incomplete Target Areas (raw number, range 0-3)	Mean Variance SD	2.37 1.14 1.07	0.07 0.14 0.38	T-test F-test	**<0.001** **<0.001**
Number of badly Placed clips (BPC) (raw number)	Mean Variance SD	2.80 6.51 2.55	0.44 0.99 0.99	T-test F-test	**<0.001** **<0.001**

8.2.2 Performance one analysis (P1)

Quantitative analysis of performance one (prior to study feedback interventions) was performed to examine for potential differences in baseline ability between Group A (information transfer feedback) and B (dialogic feedback). The results are summarised in Table 60.

A significant P1 difference was identified between Group A and B in relation to BPC (p=0.02). Group B placed statistically fewer bad clips that Group A, conferring a potential superior latent ability with this skill. In relation to the other three quantitative measures of efficiency (TTC and CIPL) and accuracy (ITA), no significant inter-group differences were identified in baseline participant performance.

With respect to variation, a significant inter-group difference was identified between Group A and B with respect to TTC and BPC in P1 (p=0.008 and 0.001 respectively). The baseline variation in these measures of performance was significantly lower in Group B compared to Group A, making the performances of Group B participants more clustered as compared to Group A participants. However, in relation to the other quantitative performance measures (CIPL and ITA), there were no significant inter-group differences in relation to intra-group variation in initial performance.

Table 54: Performance one analysis

		Group A (ITF)	Group B (DF)		P values
Time taken to Complete (TTC) (seconds)	Mean Variance SD	394.77 71407.31 267.22	316.43 25387.77 159.34	T-test F-test	0.169 **0.008**
Combined instrument path length (CIPL) (metres)	Mean Variance SD	705.39 1592.71 398.97	621.50 146364.74 382.58	T-test F-test	0.823 0.4052
Number of Incomplete Target Areas (raw number, range 0-3)	Mean Variance SD	2.26 1.26 1.13	2.37 1.14 1.07	T-test F-test	0.700 0.776
Number of badly Placed clips (BPC) (raw number)	Mean Variance SD	5.323 27.69 5.26	2.80 6.51 2.55	T-test F-test	**0.002** **0.001**

8.2.3 Time taken to complete the task (TTC)

The time taken to complete the task (TTC) for each performance (P1 – P8) was recorded for each participant and the group mean TTC at each performance was calculated and is illustrated in Table 55. TTC is a measure of efficiency of performance.

8.2.3.1 Direct performance comparison: P1 – P8

There were no significant differences in the mean TTC achieved by Group A (ITF) and B (DF) in each of the eight task performances, nor in the minimum (best) TTC achieved by each group (Table 55). The intra-group variation in Group B was significantly lower when comparing TTC in P3, P4, P5 and P6.

Table 55: Comparison of TTC performance P1 – P8

	Mean TTC (seconds)		P values	
	Group A (ITF)	**Group B (DF)**	**T-test**	**F-test**
P1	389	310	0.190	**0.008**
P2	264	308	0.298	0.331
P3	341	249	0.250	**<0.001**
P4	220	181	0.320	**0.010**
P5	192	175	0.544	**0.018**
P6	162	137	0.322	**<0.001**
P7	132	132	0.973	0.826
P8	172	134	0.107	0.106
Best TTC	107	107	0.980	0.375

8.2.3.2 Comparison of trend in performance: P1 – P8

The average performance of both groups improved between the beginning (P1) and the end (P8) of the study (Table 56). There was no significant difference between Group A and B in relation to the trend in TTC over the whole length of the study (p=0.429). However, intra-group variation in this trend was significantly different (p=0.003), with the variation in TTC trend significantly lower in Group B compared to Group A. This is illustrated by Figures 56 (TTC in Group A) and 57 (TTC in Group B). These figures plot the trend in TTC performance for each individual participant within Group A and B respectively. The decreased variation in performance in Group B is illustrated by the smaller pitch of these plots in Figure C compared to Figure B.

Table 56: Analysis of TTC trend in performance (P1 – P8) (seconds)

	Group A (ITF)	Group B (DF)		P values
Mean gradient (SD)	-34.18 (27.88)	-29.43 (15.50)	T-test	0.429
Variance	777.22	240.11	F-test	**0.003**

Figure 56: Group A (ITF) individual participant TTC, P1 – P8

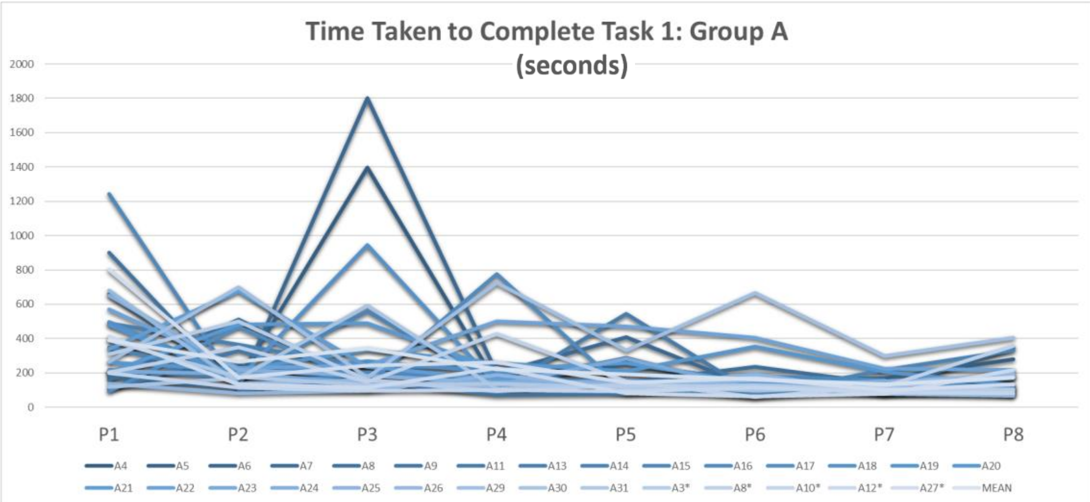
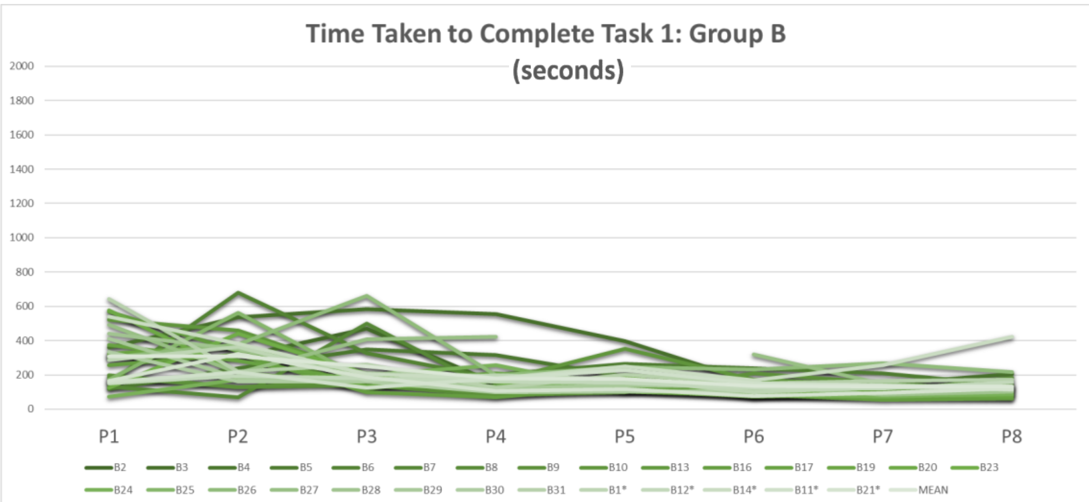


Figure 57: Group B (DF) individual participant TTC, P1 – P8



8.2.2.3 Comparison of intra-visit performance trend

No significant inter-group differences were observed in relation to the intra-visit gradients of mean TTC (Table 57). Improvement in TTC was observed in both study groups during study visits 1, 2 and 3. However, although the changes did not reach statistical significance, the mean gradient in relation to TTC in visit 2 and 3 for Group A was -28 and -30 respectively, but the mean gradient in Group B was only -3 and -4 seconds. Therefore, Group A tended towards greater intra-*visit* improvements in TTC performance.

Significant inter-group differences were observed in relation to the consistency of intra-visit change in TTC (p <0.001 for all three study visits), with the variation of TTC change less in Group B compared to Group A (Table 63).

Table 57: Comparison of intra-visit TTC performance(seconds)

		Group A (ITF)	**Group B (DF)**	**P values**	
Visit 1 (P1 – P3)	Mean (SD)	-24.17	-30.68	T-test	0.897
	Variance	58934.59	12316.28	F-test	**<0.001**
Visit 2 (P4 – P5)	Mean (SD)	-28.03	-2.81	T-test	0.435
	Variance	37631.25	6152.54	F-test	**<0.001**
Visit 3 (P6 – P7)	Mean (SD)	-30.41	-4.32	T-test	0.184
	Variance	8539.04	2186.97	F-test	**<0.001**

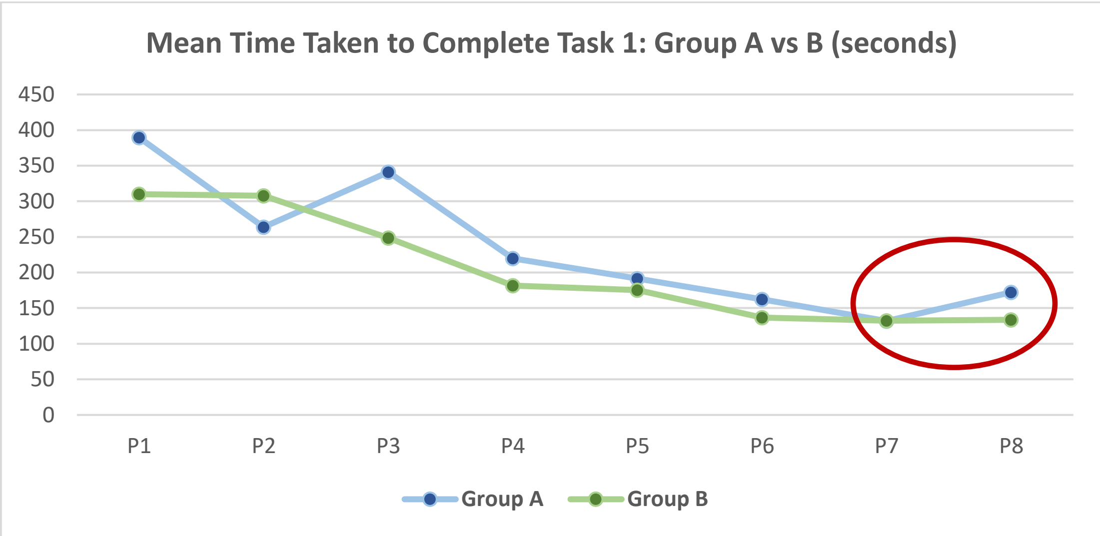
8.2.3.4 Comparison of inter-visit performance trend

No significant inter-group differences were observed in relation to the visit 1 – 2 or visit 2 – 3 changes in mean TTC performance (Table 58). Improvement in mean TTC was observed in both study groups during study visits 1 and 2. However, a significant inter-group difference was observed in relation to the *inter-visit* change in mean TTC between visit 3 and 4 (a hiatus of four weeks). In Group B, the mean gradient in TTC between the end of study visit 3 and study visit 4 was +1 but +41 for Group A (p=0.034). These results suggest that whilst participants in Group B were able to maintain their efficiency in relation to TTC after a four-week break in the study (a measure of sustained response to feedback), efficiency in Group A deteriorated. This significant difference is highlighted in Figure 58.

Table 58: Comparison of inter-visit TTC performance (seconds)

		Group A (ITF)	Group B (DF)	P values	
Visit 1 - 2 (P3 – P4)	Mean (SD) Variance	-121.38 207443.17	-67.18 16545.04	T-test F-test	0.542 **<0.001**
Visit 2 - 3 (P5 – P6)	Mean (SD) Variance	-29.28 (142) 20033.99	-45.41 (68) 4560.87	T-test F-test	0.586 **<0.001**
Visit 3 -4 (P7 – P8)	Mean (SD) Variance	+41.12 (80) 6328.75	+0.70 (51) 2630.14	T-test F-test	**0.034** **<0.001**

Figure 58: Mean TTC for P1 – P8, Group A (ITF) vs B (DF)



Similar to the observations made in intra-visit comparisons, significant inter-group differences were observed in relation to the consistency of inter-visit change in TTC (p <0.001 for all three comparisons), with the variation of TTC change less in Group B compared to Group A (Table 58).

8.2.4 Combined Instrument Path Length (CIPL)

8.2.4.1 Direct performance comparison: P1 – P8

The group mean combined instrument path length (CIPL, how far the participants needed to move both instruments in order to perform the task) in each performance (P1 – P8) was calculated and is shown in Table 59.

There was a single significant inter-group difference between Group A (ITF) and B (DF) in relation to P5 mean CIPL (p<0.001). There were no other significant differences in the mean CIPL achieved by Group A and B in each of the other seven task performances, nor in the average of the best CIPLs achieved by the participants in each group. The intra-group variation in Group B was significantly lower when comparing the CIPL of P5 and P8 (p<0.001).

Table 59: Comparison of CIPL performance P1 – P8

	Mean CIPL (metres)		**P values**	
	Group A (ITF)	**Group B (DF)**	**T-test**	**F-test**
P1	703.34	594.39	0.333	0.180
P2	424.24	554.79	0.123	0.810
P3	415.18	403.29	0.891	0.089
P4	424.24	350.46	0.347	0.619
P5	1104.21	293.22	**<0.001**	**<0.001**
P6	285.93	252.71	0.505	0.091
P7	201.21	201.75	0.980	0.675
P8	282.27	219.61	0.133	**0.045**
Best CIPL	172.48	174.00	0.916	0.459

8.2.4.2 Comparison of trend in performance: P1 – P8

The average performance of both groups improved between the beginning (P1) and the end (P8) of the study (Table 60). There was no significant difference between Groups A and B in relation to the trend in CIPL over the whole length of the study (p=0.173), nor in intra-group variation (p=0.601).

Table 60: Analysis of CIPL trend in performance (P1 – P8) (centimetres)

	Group A (ITF)	Group B (DF)		P values
Mean gradient (SD)	-41.05 (37.44)	-54.99 (41.24)	T-test	0.146
Variance	1401.62	1700.67	F-test	0.695

8.2.4.3 Comparison of intra-visit performance trend

A significant inter-group difference was observed in relation to the visit two CIPL gradient (Table 61), with Group B improving during this study (gradient -44) and Group A's performance deteriorating (gradient +659). This is related to the large rise observed in P5 CIPL in Group A. No similar trend was observed in relation to study visit 1 or 3, when Group A's improvement in mean CIPL was greater than that observed in Group B (not statistically significant).

Significant inter-group differences were observed in relation to the consistency of CIPL performance during 3 (p=0.040), with the variation of CIPL change less in Group B compared to Group A (Table 67).

Table 61: Comparison of intra-visit CIPL performance (centimetres)

		Group A (ITF)	Group B (DF)		P values
Visit 1 (P1 – P3)	Mean (SD)	-149.19	-95.55	T-test	0.425
	Variance	85168.29	42792.84	F-test	0.080
Visit 2 (P4 – P5)	Mean (SD)	+679.97	-44.00	T-test	**<0.001**
	Variance	530268.01	34830.94	F-test	**<0.001**
Visit 3 (P6 – P7)	Mean (SD)	-84.72 (184)	-50.96	T-test	0.418
	Variance	33913.92	15087.82	F-test	**0.040**

8.2.4.4 Comparison of inter-visit performance trend

No significant inter-group differences were observed in relation to the visits 1 – 2 or visits 3 – 4 changes in mean CIPL performance (Table 62). However, a significant inter-group, inter-visit difference was observed in relation to the change in mean CIPL between visit 2 – 3. In Group B, the gradient of mean change in CIPL between the end of study visit 2 and start of study visit 3 was -60 but in Group A, the gradient was -818 (p <0.001). This finding is related to the large rise observed in Group A P5 CIPL, which is incongruous with the general trend.

Significant inter-group differences were observed in relation to the consistency of inter-visit change in CIPL (p <0.001 for all three comparisons), with the variation of CIPL change less in Group B compared to Group A (Table 68).

Table 62: Comparison of inter-visit CIPL performance (centimetres)

		Group A (ITF)	Group B (DF)	P values	
Visit 1 - 2 (P3 – P4)	Mean (SD) Variance	+16.54 242342.21	-52.82 41351.50	T-test F-test	0.503 **<0.001**
Visit 2 - 3 (P5 – P6)	Mean (SD) Variance	-818.28 473858.46	-60.11 19930.22	T-test F-test	**<0.001** **<0.001**
Visit 3 -4 (P7 – P8)	Mean Variance	+78.88 25288.68	+17.86 (78) 6007.79	T-test F-test	0.069 **<0.001**

8.2.5 Incomplete target areas (ITA)

8.2.5.1 Direct performance comparison: P1 – P8

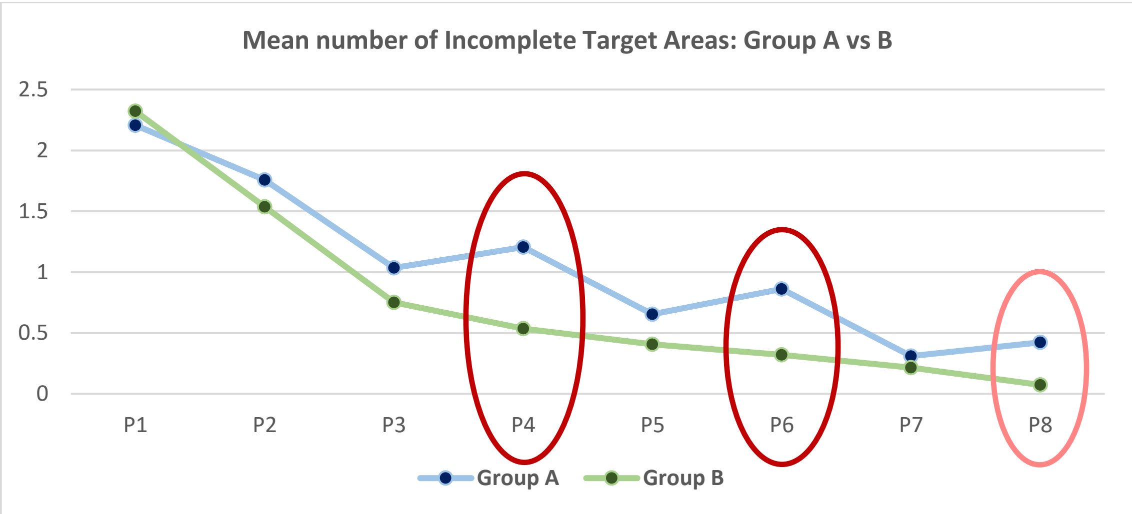
The number of incomplete target areas associated with each performance (P1 – P8) was recorded for each participant and the group mean ITA for each performance was calculated (Table 63).

There were significant differences in the mean ITA achieved by Group A (ITF) and B (DF) in P4 (p=0.018) and P6 (p=0.022), with differences approaching significance in P8 (p=0.066). In these performances, Group B achieved a significantly lower number of ITAs compared to Group A. These three performances correspond with the first task performance in study visits two, three and four respectively. The significant performances are highlighted in Figure 59. The intra-group variation in Group B was significantly lower when comparing ITA in P6 and P8 (Table 69).

Table 63: Comparison of mean ITA, performance P1 – P8

	Mean CIPL (metres)		P values	
	Group A (ITF)	Group B (DF)	T-test	F-test
P1	2.21	2.32	0.700	0.799
P2	1.76	1.54	0.500	0.842
P3	1.04	0.75	0.359	0.683
P4	1.21	0.54	0.018	0.447
P5	0.66	0.41	0.313	0.466
P6	0.86	0.32	0.022	**0.005**
P7	0.31	0.21	0.558	0.437
P8	0.42	0.07	0.066	**<0.001**
Best ITA	0.00	0.00	N/A	N/A

Figure 59: Mean ITA for P1 – P8, Group A (ITF) vs B (DF)



8.2.5.2 Comparison of trend in performance: P1 – P8

The average performance of both groups improved between the beginning (P1) and the end (P8) of the study (Table 64). There was no significant difference between Groups A and B in relation to the trend in ITA over the whole length of the study (p=0.507). Intra-group variation in this trend was significantly not different (p=0.516).

Table 64: Comparison of ITA trend in performance (P1 – P8) (seconds)

	Group A (ITF)	Group B (DF)	P values	
Mean gradient (SD) Variance	-0.25 (0.18) 0.03	-0.28(0.16) 0.02	T-test F-test	0.507 0.516

8.2.5.3 Comparison of intra-visit performance trend

No significant inter-group differences were observed in relation to the intra-visit gradients of mean ITA (Table 65). Improvement in ITA was observed in both study groups during study visits 1, 2 and 3. This difference approached significance in visit 3 (p=0.053). It is interesting to note that the mean improvement in ITA was greater in visit 2 and 3 in Group A compared to Group B. Therefore, Group A tended towards greater intra-visit improvements in ITA performance.

Significant inter-group differences were observed in relation to the consistency of intra-visit change in ITA in relation to visit 3 (p <0.001), with the variation of ITA change less in Group B compared to Group A (Table 65).

Table 65: Comparison of intra-visit ITA performance

		Group A (ITF)	Group B (DF)	P values	
Visit 1 (P1 – P3)	Mean (SD) Variance	-0.53 (0.74) 0.55	-0.79 (0.69) 0.47	T-test F-test	0.191 0.681
Visit 2 (P4 – P5)	Mean (SD) Variance	-0.55 (1.43) 2.04	-0.15 (1.13) 1.28	T-test F-test	0.246 0.239
Visit 3 (P6 – P7)	Mean (SD) Variance	-0.55 (1.12) 1.26	-0.11 (0.42) 0.17	T-test F-test	0.053 **<0.001**

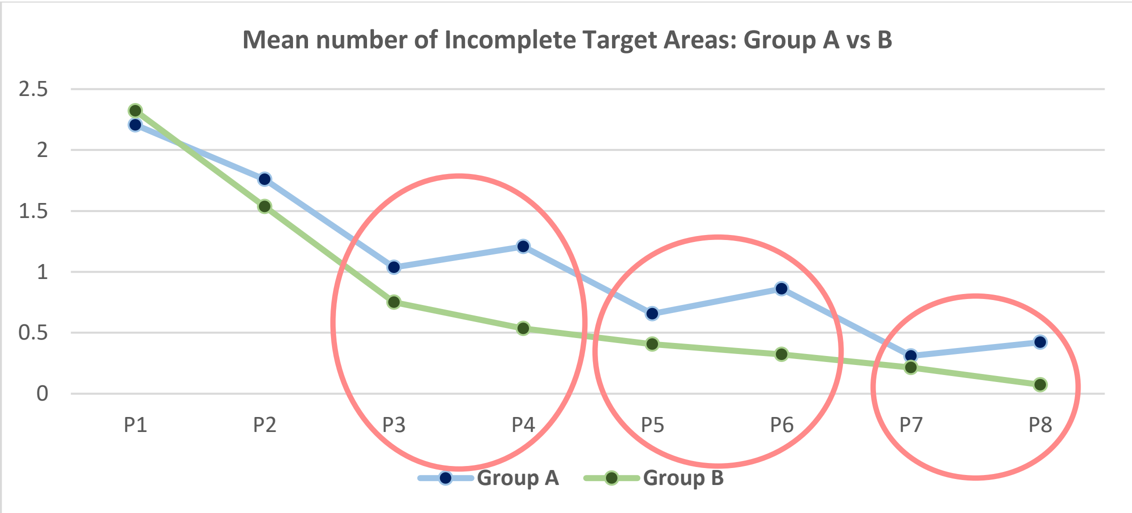
8.2.5.4 Comparison of inter-visit performance trend

No significant inter-group differences were observed in relation to inter-visit changes in mean ITA performance (Table 66). However, an interesting inter-group, inter-visit trend is apparent. In Group A, number of ITAs increased between study visits (as denoted by positive gradients). However, in Group B, the number of ITAs decreased between study visits. This pattern is highlighted in Figure 60. These results suggest that whilst participants in Group B were able to maintain and even improve their accuracy in relation to ITA after breaks between study visits (a measure of sustained response to feedback), accuracy in Group A deteriorated.

Table 66: Comparison of inter-visit ITA performance

		Group A (ITF)	Group B (DF)	P values	
Visit 1 - 2 (P3 – P4)	Mean (SD)	+0.11 (1.55)	-0.21 (1.23)	T-test	0.393
	Variance	2.40	1.51	F-test	0.236
Visit 2 - 3 (P5 – P6)	Mean (SD)	+0.21 (1.42)	-0.11 (0.85)	T-test	0.311
	Variance	2.03	0.72	F-test	**0.009**
Visit 3 -4 (P7 – P8)	Mean (SD)	+0.08 (1.13)	-0.15 (0.72)	T-test	0.393
	Variance	1.27	0.52	F-test	**0.026**

Figure 60: Mean ITA for P1 – P8, Group A (ITF) vs B (DF)



8.2.6 Badly placed clips (BPC)

8.2.6.1 Direct performance comparison: P1 – P8

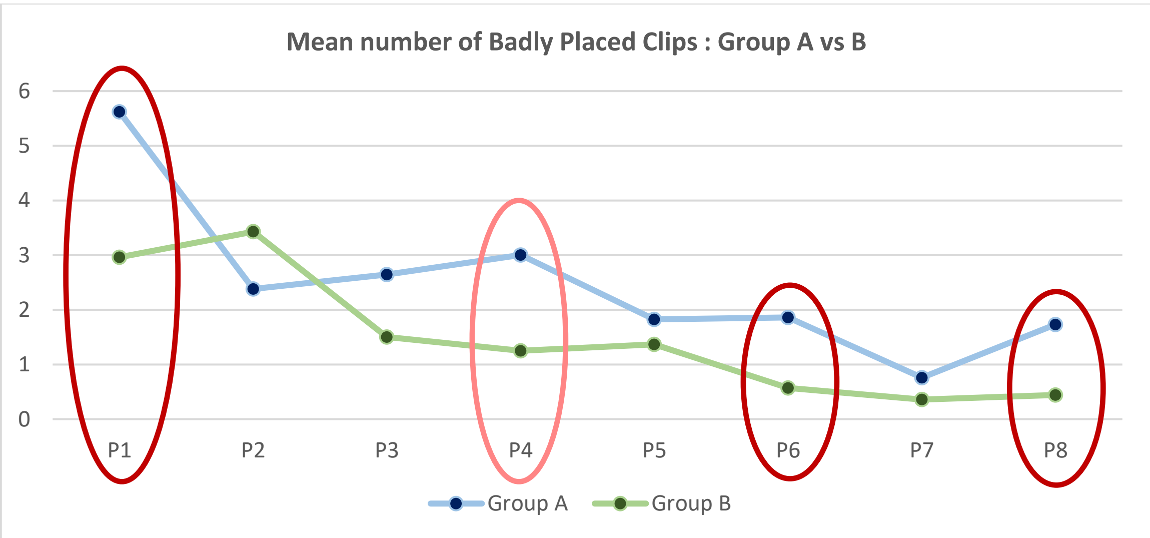
The number of badly placed clips (BPC) inserted during each task performance (P1 – P8) was recorded for each participant and the group mean BPC at each performance was calculated and is illustrated in Table 67 and Figure 61.

There was a significant inter-group difference in the mean BPC achieved by Group A (ITF) and B (DF) in P1 (p=0.020), P6 (p=0.041) and P8 (p=0.009). The difference at P4 approached significance (p=0.075). The significant performances are highlighted in Figure 61. In relation to these performances, Group B placed fewer badly placed clips than Group A. These performances correspond with the first task performance in each study visit. The significant difference in the baseline, pre-intervention performance (P1) highlights a potential confounding factor. However, the inter-group difference was not sustained during study visit one. Significant inter-group differences were observed in relation to the consistency of intra-visit change in BPC; variation in Group B was significantly lower in P1, P4, P6 and P8 (p<0.001).

Table 67: Comparison of BPC performance P1 – P8

	Mean BPC (raw number)		**P values**	
	Group A (ITF)	**Group B (DF)**	**T-test**	**F-test**
P1	5.62	2.96	**0.020**	**<0.001**
P2	2.38	3.43	0.300	0.230
P3	2.64	1.50	0.144	0.038
P4	3.00	1.25	0.075	**<0.001**
P5	1.83	1.37	0.414	0.104
P6	1.86	0.57	**0.041**	**<0.001**
P7	0.76	0.36	0.101	0.003
P8	1.73	0.44	**0.009**	**<0.001**
Best BPC	0.03	0.04	0.980	0.979

Figure 61: Mean BPC for P1 – P8, Group A (ITF) vs B (DF)



8.2.6.2 Comparison of trend in performance: P1 – P8

The average performance of both groups improved between the beginning (P1) and the end (P8) of the study (Table 68). There was no significant difference between Groups A and B in relation to the trend in BPC over the whole length of the study (p=0.674). There was no significant difference in relation to intra-group variation (p=0.147).

Table 68: Analysis of BPC trend in performance (P1 – P8) (seconds)

	Group A (ITF)	Group B (DF)		P values
Mean gradient (SD)	-0.48 (0.54)	-0.43 (0.41)	T-test	0.674
Variance	0.30	0.17	F-test	0.147

8.2.6.3 Comparison of intra-visit performance trend

No significant inter-group differences were observed in relation to the intra-visit gradients of mean BPC (Table 69). However, although the changes did not reach statistical significance, it is interesting to note that the mean change in BPC showed greater intra-visit improvement in Group A as compared to Group B.

Significant inter-group differences were observed in relation to the consistency of intra-visit change in BPC (P <0.001 for all three study visits), with the variation of BPC change less in Group B compared to Group A (Table 69).

Table 69: Comparison of intra-visit BPC performance

		Group A (ITF)	Group B (DF)		P values
Visit 1 (P1 – P3)	Mean (SD)	-1.57 (3.13)	-0.73 (1.83)	T-test	0.227
	Variance	9.77	3.34	F-test	**0.007**
Visit 2 (P4 – P5)	Mean (SD)	-1.17 (5.30)	+0.11	T-test	0.250
	Variance	28.08	6.49	F-test	**<0.001**
Visit 3 (P6 – P7)	Mean (SD)	-1.10 (3.38)	-0.21 (1.10)	T-test	0.188
	Variance	11.45	1.21	F-test	**<0.001**

8.2.6.4 Comparison of inter-visit performance trend

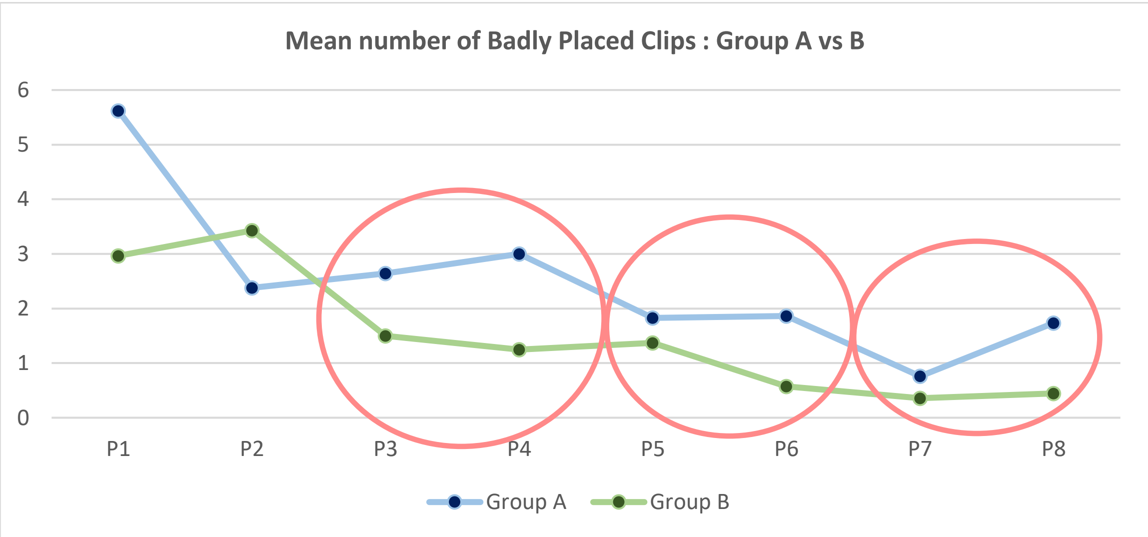
No significant inter-group differences were observed in relation to inter-visit changes in mean BPC performance (Table 70). However, an interesting inter-group trend is apparent. In Group A, number of BPCs *increased* between study visits (as denoted by positive gradients). However, in Group B, the number of BPCs decreased between study visits 1 – 2 and 2 – 3, with a smaller positive gradient between visit 3 – 4 (+0.07 versus +1.00). This pattern is highlighted in Figure 62.

Significant inter-group differences were observed in relation to the consistency of inter-visit change in BPC ($p \leq 0.001$ for all three comparisons), with the variation of BPC change less in Group B compared to Group A (Table 70).

Table 70: Comparison of inter-visit BPC performance

		Group A (ITF)	Group B (DF)		P values
Visit 1 - 2 (P3 – P4)	Mean (SD) Variance	+0.36 (6.14) 37.65	-0.25 (3.17) 10.05	T-test F-test	0.644 **0.001**
Visit 2 - 3 (P5 – P6)	Mean (SD) Variance	+0.03 (4.15) 17.25	-0.81 (2.20) 4.85	T-test F-test	0.339 **0.001**
Visit 3 -4 (P7 – P8)	Mean (SD) Variance	+1.00 (2.56) 6.50	+0.07 0.69	T-test F-test	0.089 **<0.001**

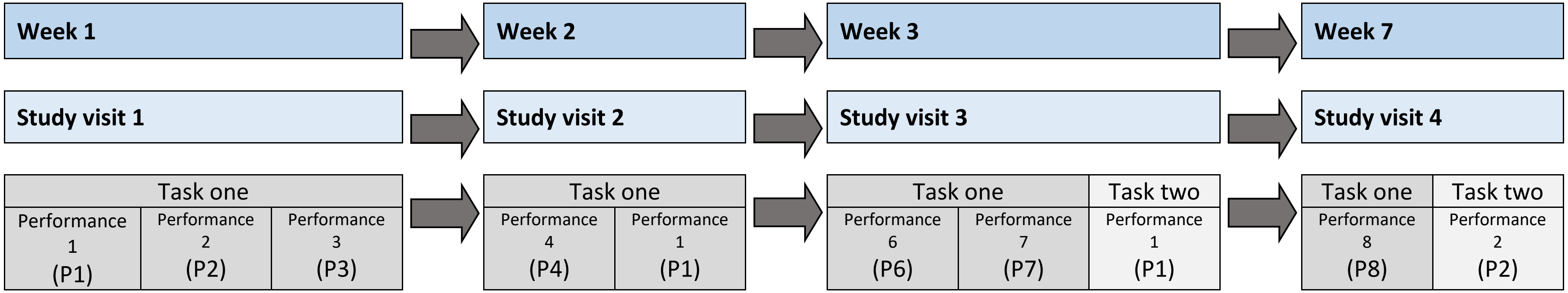
Figure 62: Mean BPC for P1 – P8, Group A (ITF) vs B (DF)



8.3 TASK TWO PERFORMANCE ANALYSIS (VESSEL SECTIONING)

Figure 55 is repeated here to illustrate the timing of study visits and task two performances and acts as a reference during interpretation of this results section.

Figure 55: Chart illustrating timing of study visits and task one and task two performances



Participants did not receive feedback in relation to task two. Participant performance of this cross-over task was used to assess the ability of participants to apply the simulated laparoscopic skills garnered from execution of task one to another simulated laparoscopic task. Participants performed the task a total of twice; the first time during visit 3 (P1) and then second time four weeks later during visit 4 (P2).

8.3.1 Intra-group performance improvement (P2 vs P1)

Only Group B demonstrated a significant improvement in performance in respect to TTC ($p=0.008$) over the course of the study (P2 vs P1 performance, Table 71). This suggests that engagement in a dialogic model of feedback is associated with improved efficiency of performance of psychomotor tasks, even in the absence of direct feedback in relation to this task. No other statistically significant differences were seen in relation to CIPL, or BPC and ITA. There were no significant quantitative changes in performance in relation to the cross-over task observed in Group A.

Table 71: Intra-group performance analysis (P2 vs P1)

		Group A (P1)	Group A (P2)		P values
Time taken to Complete (TTC) (seconds)	Mean Variance SD	288.24 8086.32 89.92	247.69 10362.14 101.79	T-test F-test	288.24 8086.32 89.92
Combined instrument path length (CIPL) (metres)	Mean Variance SD	432.14 50049.77 223.72	433.81 59292.62 243.50	T-test F-test	432.14 50049.77 223.72
Frequency of abdominal wall damage (AWD) (raw number)	Mean Variance SD	7.41 111.35 10.55	6.81 37.16 6.10	T-test F-test	7.41 111.35 10.55
Maximum depth of abdominal wall damage (MDD) (cm)	Mean Variance SD	8.58 97.93 9.90	5.27 21.86 4.68	T-test F-test	8.58 97.93 9.90
		Group A (P1)	**Group A (P2)**		**P values**
Time taken to Complete (TTC) (seconds)	Mean Variance SD	231.36 2998.16 54.76	195.11 1598.91 39.99	T-test F-test	231.36 2998.16 54.76
Combined instrument path length (CIPL) (metres)	Mean Variance SD	357.61 18626.10 136.48	357.52 11671.81 108.04	T-test F-test	357.61 18626.10 136.48
Frequency of abdominal wall damage (AWD) (raw number)	Mean Variance SD	4.57 29.46 5.43	5.33 27.85 5.28	T-test F-test	4.57 29.46 5.43
Maximum depth of abdominal wall damage (MDD) (cm)	Mean Variance SD	4.34 24.72 4.97	3.67 8.26 2.87	T-test F-test	4.34 24.72 4.97

8.3.2 Time taken to complete the task (TTC)

8.3.2.1 Direct performance comparison: P1 – P2

The time taken to complete the task (TTC) for each performance (P1 – P2) was recorded for each participant and the group mean TTC at each performance was calculated. Table 72 illustrates that there was a significant difference in the mean TTC achieved by Group A (ITF) and B (DF) in both task two performances (P1 p=0.007; P2 p=0.022). The intra-group variation in Group B (DF) was significantly lower when comparing TTC in both P1 and P2 (p <0.001).

Table 72: Comparison of TTC performance P1 – P2

	Mean TTC (seconds)		**P value**	
	Group A (ITF)	**Group B (DF)**	**T-test**	**F-test**
P1	288.24	231.36	**0.007**	**0.012**
P2	247.69	195.11	**0.022**	**<0.001**

8.3.2.2 Comparison of trend in performance: P1 – P2

The average performance of both groups improved between the first and second performance of task two (Table 73). There was no significant difference between Groups A and B in relation to the trend in TTC change. However, intra-group variation in this trend was significantly different (p<0.001), with the variation in TTC trend significantly lower in Group B compared to Group A.

Table 73: Analysis of TTC trend in performance (P1 – P2) (seconds)

	Group A (ITF)	**Group B (DF)**		**P value**
Mean gradient	-43.19 (111)	-37.89 (59)	T-test	0.832
Variance	12448.56	3479.18	F-test	**<0.001**

8.3.3 Combined Instrument Path Length (CIPL)

8.3.3.1 Direct performance comparison: P1 – P2

The combined instrument path length (CIPL) for each performance (P1 – P2) was recorded for each participant and the group mean CIPL for each performance was calculated. Table 74 illustrates that there was no significant difference in the mean CIPL achieved by Group A (ITF) and B (DF) in both task two performances. The intra-group variation in Group B was significantly lower when comparing CIPL in both P1 (p=0.012) and P2 (p <0.001).

Table 74: Comparison of CIPL performance P1 – P2

	Mean TTC (seconds)			P value
	Group A (ITF)	**Group B (DF)**	**T-test**	**F-test**
P1	432.14	357.61	0.141	**0.012**
P2	433.81	357.52	0.160	**<0.001**

8.3.3.2 Comparison of trend in performance: P1 – P2

The average performance of both groups was extremely consistent between the first and second performance of task two (Table 75). There was no significant difference between Groups A and B in relation to the trend in TTC change. However, intra-group variation in this trend was significantly different (p=0.001), with the variation in TTC trend significantly lower in Group B compared to Group A.

Table 75: Analysis of CIPL trend in performance (P1 – P2) (centimetres)

	Group A (ITF)	Group B (DF)		P value
Mean gradient	-6.38 (266)	-3.33 (141)	T-test	0.997
Variance	71006.33	19822.92	F-test	**0.001**

8.3.4 Frequency of abdominal wall damage (AWD)

8.3.4.1 Direct performance comparison: P1 – P2

The frequency of occasions that the simulated laparoscopic instruments damaged the abdominal wall (AWD) during each performance (P1 – P2) was recorded for each participant and the group mean AWD for each performance was calculated. Table 76 illustrates that there was no significant difference in the mean AWD achieved by Group A (ITF) and B (DF) in both task two performances. The intra-group variation in Group B was significantly lower when comparing AWD in P1 (p <0.001) but not P2.

Table 76: Comparison of AWD performance P1 – P2

	Mean TTC (seconds)			P value
	Group A (ITF)	**Group B (DF)**	**T-test**	**F-test**
P1	7.414	4.571	0.214	**<0.001**
P2	6.808	5.333	0.361	0.468

8.3.4.2 Comparison of trend in performance: P1 – P2

The average performance of both groups was extremely consistent between the first and second performance of task two (Table 77). There was no significant difference between Groups A and B in relation to the trend in AWD change. However, intra-group variation in this trend was significantly different (p=0.006), with the variation in AWD trend significantly lower in Group B compared to Group A.

Table 77: Analysis of AWD trend in performance (P1 – P2)

	Group A (ITF)	**Group B (DF)**		**P value**
Mean gradient	-0.96 (10.99)	+0.89 (6.38)	T-test	0.597
Variance	120.84	40.64	F-test	**0.006**

8.3.5 Maximum depth of abdominal wall damage (MDD)

8.3.5.1 Direct performance comparison: P1 and P2

The maximum depth of the damage sustained to the abdominal wall from the simulated laparoscopic instruments (MDD) during each performance (P1 – P2) was recorded for each participant and the group mean MDD for each performance was calculated and illustrated in Table 78.

There was a significant difference in the mean AWD achieved by Group A (ITF) and B (DF) in task two P1 (p=0.049) but not P2. Group B achieved a significantly smaller maximum depth of abdominal wall damage compared to Group A (4.34 vs 8.58 mm respectively). The intra-group variation in Group B was significantly lower when comparing AWD in P1 (p <0.001) and P2 (p=0.016).

Table 78: Comparison of MDD performance P1 and P2

	Mean TTC (seconds)		P value	
	Group A (ITF)	Group B (DF)	T-test	F-test
P1	8.58	4.34	**0.049**	**<0.001**
P2	5.27	3.67	0.150	**0.016**

8.3.5.2 Comparison of trend in performance: P1 – P2

There was no significant difference between Groups A and B in relation to the trend in MDD change between P1 and P2 (Table 79). However, intra-group variation in this trend was significantly different (p<0.001), with the variation in MDD trend significantly lower in Group B compared to Group A.

Table 79: Analysis of MDD trend in performance (P1 – P2) (centimetres)

	Group A (ITF)	Group B (DF)	P value	
Mean gradient	-3.31 (9.99)	-0.49 (5.21)	T-test	0.200
Variance	99.98	27.17	F-test	**<0.001**

8.4 POST-STUDY PARTICIPANT CONFIDENCE & STUDY SATISFACTION ANALYSIS

The post-study participant self-confidence in relation to performing laparoscopic tasks and confidence in the tutor (researcher) in providing useful feedback was measured at the end of this study (Table 80). As with the pre-study questionnaire, participants were asked to score each of these factors on a Likert scale between 1 (lowest confidence) and 5 (highest confidence). Participants were asked to score their satisfaction with participation in the study on a Likert scale between 1 (conferring the lowest satisfaction) and 5 (conferring maximum satisfaction).

Neither the post-study participant self-confidence nor confidence in tutor was significantly different between groups. Both groups highly rated their confidence in the tutor and satisfaction with participation in the study.

Table 80: Post-study confidence and satisfaction analysis of participants

		Group A (ITF)	Group B (DF)	P value
Post-study self-confidence (range 0 – 5)	Mean SD	3.46 0.84	3.31 0.76	0.471
Post-study tutor-confidence (range 0 – 5)	Mean SD	4.75 0.44	4.86 0.35	0.294
Post-study feedback satisfaction (range 0 – 5)	Mean SD	4.5357 0.5762	4.5517 0.5061	0.912

8.5 SUMMARY OF RESULTS

Adherence to study protocol in relation to timing of study visits was accurate and inter-visit intervals were similar between groups. There were no significant inter-group differences in relation to feedback volume. There were no significant inter-group differences in relation to the pre-study experience, pre-study self- or tutor-confidence, hand dominance or age. This supports the random allocation method used in this study. However, there was a significant difference in the proportion of male and female participants allocated to the groups (p=0.049). There were no significant inter-group differences in relation to post-study participant self-confidence, confidence in the tutor or satisfaction with study participation.

8.5.1 Task one (vessel ligation)

Table 81 summarises the quantitative findings in relation to task one inter-group analysis of performance and variation in performance. It illustrates the statistically significant findings in relation to task performance and task variation and notable non-statistically significant trends observed.

Table 81: Task one: Summary of statistical analysis

	MEASURES OF EFFICIENCY (TTC/CIPL)			MEASURES OF ACCURACY (ITA/BPC)		
	PERFORMANCE		VARIANCE	PERFORMANCE		VARIANCE
	Statistically significant findings	Non-statistically significant trends	Statistically significant findings	Statistically significant findings	Non-statistically significant trends	Statistically significant findings
P1			TTC (B)	BPC (B)		BPC (B)
INTRA-GROUP PERFORMANCE Group A (P8 vs P1)	TTC CIPL		TTC CIPL	ITA BPC		BPC
INTRA-GROUP PERFORMANCE Group B (P8 vs P1)	TTC CIPL		TTC CIPL	ITA BPC		ITA BPC
INTER-GROUP Full study course (P1 – 8)		TTC & CIPL: Both groups improved of course of study			ITA & BPC: Both groups improved of course of study	ITA P6 & 8 (B)
INTER-GROUP Individual Performances (P1 – 8)	CIPL P5 (B)		TTC P3, 4, 5 & 6 (B) CIPL P5 (B)	ITA P4 & 6 (B) BPC P6 & 8 (B)	ITA P8 (B) BPC P4 (B)	ITA P6 & 8 (B) BPC P4, 6 & 8 (B)
INTER-GROUP Intra-visit 1, 2 AND 3	CIPL visit 2 (B)	TTC visit 2 & 3 (A)	TTC visit 1, 2 & 3 (B) CIPL visit 2 & 3 (B)		ITA Visit 2 & 3 (A) BPC visit 1, 2 & 3 (A)	ITA Visit 3 (B) BPC visit 1, 2 & 3 (B)
INTER-GROUP Inter-visit (1-2, 2-3, 3-4)	TTC visit 3-4 (B) CIPL visit 2-3 (B)	CIPL visit 1-2 & 3-4 (B)	TTC visit 1, 2 & 3 (B) CIPL visit 1, 2 & 3 (B)		ITA visit 1-2, 2-3 & 3-4 (B) BPC visit 1-2, 2-3 & 3-4 (B)	ITA visit 2-3 & 3-4 (B) BPC visit 1-2, 2-3 & 3-4 (B)
KEY	The study group with the significantly better performance or significantly lower variation is indicated by (A) for Group A (information transfer feedback) or (B) for Group B (dialogic feedback). The performance relating to significant difference or important trends is indicated by number (P1-8 referring to performance 1 – 8). TTC = Time taken to complete task; CIPL = Combined instrument path length; BPC = Number of badly placed clips; ITA = Number of incomplete target areas.					

Both study groups demonstrated significant improvement in performance with respect to all four measures of efficiency (time taken to complete, TTC, and combined instrument path length, CIPL) and accuracy (number of badly placed clips, BPC, and incomplete target areas, ITA) over the course of the study (P8 vs P1 performance, Table 81). This suggests that both information-transfer and dialogic models of feedback are associated with improved performance of psychomotor tasks. Both groups also exhibited less intra-group variation during P8 as compared to P1, suggesting an association between both models of feedback and improving consistency of performance.

Despite participant randomisation, a significant difference was noted in (pre-feedback intervention) performance one (P1) in relation to the variation of TTC (lower in Group B) and the variation of and performance in BPC (lower and better respectively in Group B). Over the full course of the study, both study groups improved their performance in relation to all four quantitative measures of performance.

When comparing individual task performances, Group B (DF) performed better than Group A (ITF) in relation to both measures of accuracy (ITA and BPC) at several points, corresponding to more accurate performances at the start of study visits (P4, 6 & 8).

Intra-visit analysis shows that Group A tended towards greater improvements in performance during study visits. Group A tended towards greater efficiency (TTC) in visit 2 and 3. Group A tended to be more accurate in relation to BPC in visit 1, 2 and 3 and ITA in visit 2 & 3 when compared with Group B performance. These findings did not achieve statistical significance.

Inter-visit analysis illustrates that Group B performed significantly better after the break *between* study visits. Group B displayed evidence of greater efficiency (TTC and CIPL), with both statistically significant differences and other non-significant notable trends in relation to inter-visit changes in performance.

Throughout the study, Group B showed evidence of significantly reduced variation in performance as compared to Group A.

8.5.2 Task two (vessel sectioning)

Table 82 summarises the quantitative findings in relation to task two inter-group analysis of performance and variation in performance. It illustrates the statistically significant findings in relation to task performance and task variation and notable non-statistically significant trends observed.

Table 82: Task two: Summary of statistical analysis

	MEASURES OF EFFICIENCY (TTC/CIPL)			**MEASURES OF ACCURACY (AWD/MDD)**		
	PERFORMANCE		**VARIANCE**	**PERFORMANCE**		**VARIANCE**
	Statistically significant findings	**Non-statistically significant trends**	**Statistically significant findings**	**Statistically significant findings**	**Non-statistically significant trends**	**Statistically significant findings**
INTRA-GROUP PERFORMANCE Group A						
INTRA-GROUP PERFORMANCE Group B	TTC					
INTER-GROUP Full study course / Inter-visit (P1 – 2)			TTC (B) CIPL (B)			AWD (B) MDD (B)
INTER-GROUP Individual performances (P1 – 2)	TTC P1 & P2 (B)	CIPL P1 & P2 (B)	TTC P1 & P2 (B) CIPL P1 & P2 (B)	MDD P1 (B)	AWD P1 (B)	AWD P1 (B) MDD P1 & P2 (B)
KEY	The study group with the significantly better performance or significantly lower variation is indicated by (A) for Group A (information transfer feedback) or (B) for Group B (dialogic feedback). The performance relating to significant difference or important trends is indicated by number (P1-8 referring to performance 1 – 8). TTC = Time taken to complete task; CIPL = Combined instrument path length; AWD = Frequency of abdominal wall damage; MDD = Maximum depth of abdominal wall damage.					

Group B demonstrated a significant improvement in performance in respect to TTC (p=0.008) over the course of the study (P2 vs P1 performance, Table 88). This suggests that engagement in a dialogic model of feedback is associated with improved efficiency of performance of psychomotor tasks, even in the absence of direct feedback in relation to this task. No other statistically significant differences were seen in relation to CIPL, or BPC and ITA. There were no significant quantitative changes in performance in relation to the cross-over task observed in Group A.

At all points during the study, Group B performed better on average than Group A. No statistically significant inter-group differences were observed in relation to accuracy or efficiency.

Group B were significantly more efficient than Group A in relation to TTC in both task performances. This trend was also observed in relation to CIPL. Group B was also significantly more accurate than Group A in P1 in relation to MDD. Again, this trend was also observed in relation to AWD.

CHAPTER 9: QUALITATIVE RESULTS

9.1 SEMI-STRUCTURED INTERVIEW METHODS

Semi-structured interviews were conducted in order to explore learner feedback experiences and to explore how feedback mediated learning. The following ideas were relevant to the research questions and served to orient the analysis:

- Sources of feedback and the role adopted by the learner during feedback
- Evidence of learner self-regulation
- Evidence of and ideas relating to the use of goals
- Changes in learner perceptions and ideas relating to future learning

9.1.1 Interview questions

The semi-structured interview comprised of nine core questions (Box 16), with additional explorative discussion of the ideas expressed. The process of constructing questions suitable for the interview was an iterative process, which occurred throughout the earlier stages of research (literature review and pilot studies). Their purpose was to allow exploration of participant ideas in relation to feedback and their learning experience during the study (addressed directly in questions one, two, seven, eight and nine). As my understanding of the subject matter matured, additional questions were formed to explore participant ideas in relation to transferability of learning in relation to the different feedback models (question three), ideas relating to the use of goals (a key feature of but not necessarily exclusive to self-regulation) (questions four and five), and participants ideas in relation to assessing progress, which is a key purpose of feedback, as supported by the literature review (question six).

Furthermore, several of the questions served another purpose in relation to triangulation of data. Questions seven, addressing participant perspective in relation to the source of feedback, could support or contradict the intended characteristics of the different feedback models, with the information transfer model intended to be tutor driven with the learner adopting a passive role, and the dialogic feedback model purporting to encourage tutor facilitation with learner adopting an active role and contributing content. Question eight and nine addressed the acceptability of the models of feedback to the learner and could serve to support or contradict quantitative data in relation to participant satisfaction.

It may have been possible to truncate this process by developing a finalised framework concurrently with the participant interview process, requiring fewer participants to be interviewed should saturation of data be observed. Concurrent data acquisition and analysis would also have afforded the opportunity to pilot interview questions. This was not undertaken and the limitations this causes in relation the construction of the semi-structured interview is acknowledged. However, this process was by its nature an exploratory one and with no previously published work addressing the subject matter of this study, of the learner experience in the context of dialogic feedback, it would be difficult to know when saturation of data (learner ideas and perceptions) had been reached.

Furthermore, one of the strengths of this mixed methods study and the inclusion of the quantitative approach included not just the appreciation of learner perspectives and ideas, but the inter-groups trends observed in relation to these ideas. It would not be possible to explore any learner perspectives and ideas that might generally be associated with one feedback model or another without interviewing the entire cohort. Whilst individual ideas were very valuable in contributing to the discourse relating to learner experience, inter-group trends in these ideas might also be valuable when hypothesising about the interaction between feedback model and learner experience.

These inter-group trends can be highlighted by expressing the frequency with which ideas were expressed by participants in both information transfer and dialogic feedback study groups. Whilst the use of numbers within qualitative research and analysis is contentious, there are sources which advocate its potential helpfulness (Pope et al, 2000; Sandelowski, 2001; Maxwell, 2010). Pope et al (2000) acknowledged that 'simple counts' might be useful in summary of analysis. Maxwell (2010) argued that qualitative researchers often used quantitative terms such as 'sometimes', 'often', or 'typically' when reporting and discussing results, and that to fail to explicitly acknowledge the numerical basis on which these conclusions were made, was an error and weakened qualitative research rather than strengthened it. Furthermore, he argued that *'providing numerical data about the distribution of observations, or the number of instances of a particular type of event or statement, helps to deal with potential challenges to these conclusions'* (p. 478). Stating the case most strongly for the place of numbers in qualitative research, Sandelowski (2001) suggested that qualitative research should not be regarded as 'antinumber', and that numbers may be helpful in deriving meaning from qualitative data, although context for the data remained paramount.

Therefore, in addition to the exploration of the ideas and perceptions expressed by the participants, via the devised thematic analysis coding framework (Figure 63), the frequency of the expression of these ideas by participants is included in the qualitative analysis, as this might serve to indicate the strength of association between ideas and participants within each study group.

The interviews were audio recorded and transcribed prior to thematic analysis and comparison of Groups A (information transfer feedback) and B (dialogic feedback).

Box 16: Semi-structured participant interview questions (visit four)

1 Having been involved in the study, what do you think the purpose of feedback is?

2 How did the feedback process in this project help you learn?

3 Having participated in the study, have you taken anything away that might help you learn in future?

4 How important is setting aims or goals when performing a task?

5 What attributes does a useful aim or goal have? Give me an example.

6 During the study, how did you assess progress? (Group A: Did you use the global rating scale?)

7 Who or what was the main source of feedback during the study? Were there any other sources?

8 Did you enjoy the feedback sessions? And why?

9 Was there anything you found difficult or unenjoyable about the feedback?

(Group A: I'm aware I did a lot of the talking; would it have been helpful for you to speak more?)

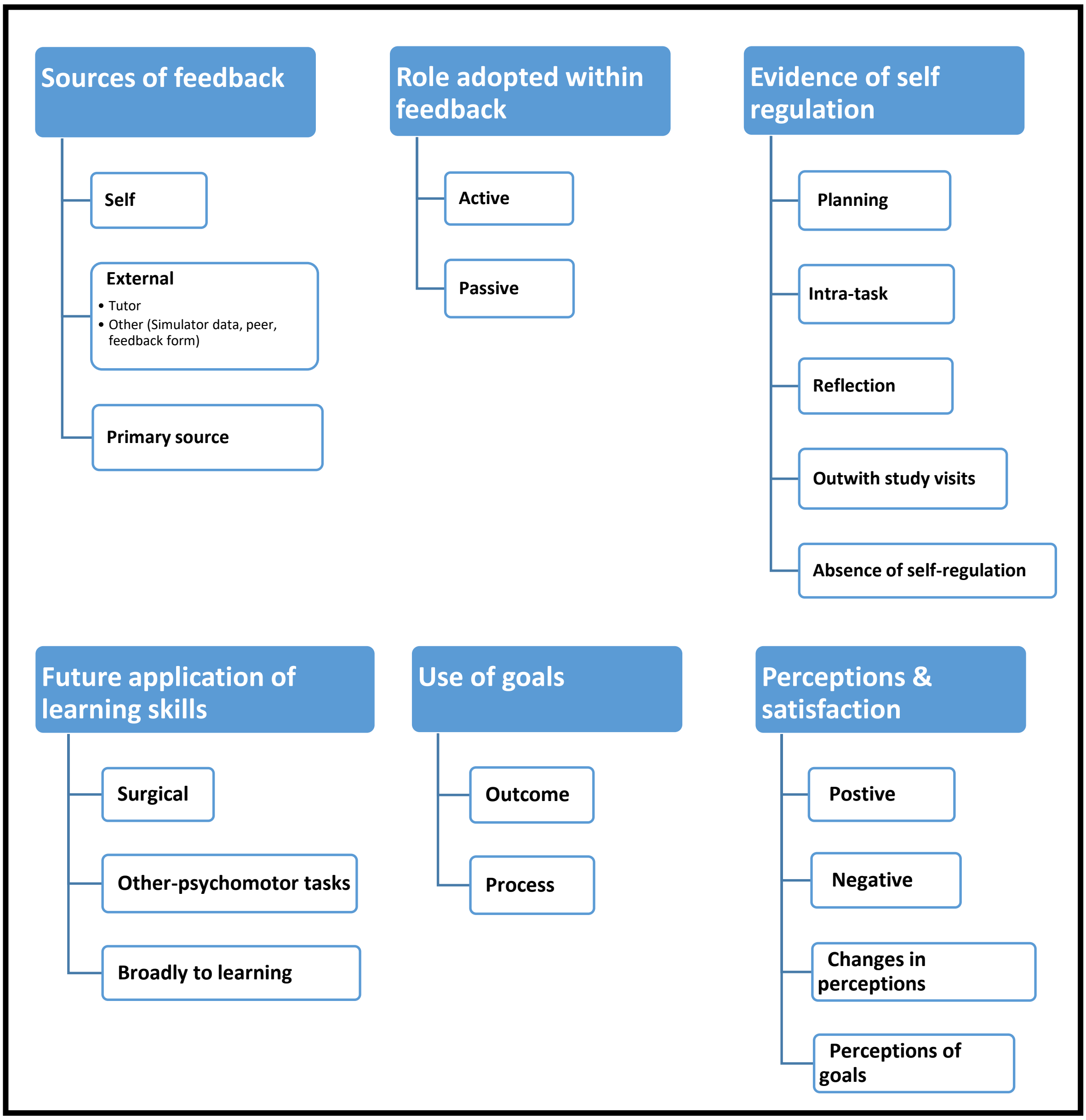
9.1.2 Participant completion

Fifty-six of the 61 participants (92%) involved in the study completed the semi-structured interview in visit four/week seven. Of the five participants unable to attend the final study visit, three were in Group A and two were in Group B. Participant demographics are detailed in the file 'Full study raw data' (sheet 'Demographics') available on the accompanying data disc.

9.1.3 Data analysis

Utilising a grounded theory methodology, a thematic data analysis approach was used to analyse the data (Richie & Spencer, 1994). The final coding framework used for thematic analysis was compiled via a number of iterative stages. As the primary researcher, I recorded my perceptions in relation to interviews as they were conducted and these observations were discussed with supervisors and related to this study's research questions. The transcribed participant interviews were read and an initial coding framework was constructed using NVivo 11 Pro software package. This preliminary framework was used as a trial and 10 participant interviews were analysed and coded. Following researcher-supervisor review of this coding, the final coding framework was constructed and used in thematic analysis of all 56 interview transcriptions(Figure 63). When transcription errors were encountered during analysis of the quantitative data (usually denoted by missing words, confusion caused by two voices speaking at once, or incongruous words) the original audio recordings were revisited and corrections were made. This increased the accuracy of transcription but also the researcher familiarity with the interview contant, promoting research validity.

Figure 63: Final thematic analysis coding framework



9.2 FINDINGS

Six main themes were explored: sources of feedback identified, the role adopted by the learner (participant) during feedback, evidence of learner self-regulation, evidence of future application of learning skills, the use of goals, and ideas expressed regarding perceptions of feedback and satisfaction. Nine hundred and sixty individual references were generated across the six broad themes and 20 coding nodes.

9.2.1 Learner roles & sources of feedback

9.2.1.1 Participant roles

Participants in Group A (information transfer feedback) were more likely to express ideas that portrayed a passive role during feedback in comparison to Group B (dialogic feedback; 27 participants, 59 items vs eight participants, ten items). Conversely, participants in Group B expressed ideas relating to contributing an active role during feedback sessions more commonly than those in Group A (23 participants and 77 separate coded items, vs six participants and seven items).

Participants in Group A cited adoption of a passive role within feedback as positive in the context of more valuable tutor expertise and instruction, with the importance of expert knowledge of the tutor promoted above the self-generation of learner thought:
"It was nice to hear it from you rather than me trying to make stuff up. Or trying to work out something where I might not necessarily know what I'm talking about. I'd rather hear it from an expert who knows what they're doing." (A24, male, 22)

Unsurprisingly, with the learner adopting a passive role within feedback, feelings of self-efficacy were not promoted. Group A participants placed limitations of the value of their input during feedback, based upon their lack of experience with the psychomotor task, demoting their agency within learning:
"And so I thought, 'Would it have been more helpful if I'd chipped in with stuff?'. I'm not sure that it would've. Because like I say, I think I improved most, if anything from just doing the task more often…. I don't have any experience using stuff like this so to get your opinion and advice it was probably more useful than if I was talking." (A8, male, 25)

However, the limitations of a feedback model that encouraged the learner to act as recipient did not go unnoticed by all participants in this arm of the study (five participants, seven coded items). These observations related to the repetitive nature of highly-structured feedback, discouragement of engagement in feedback and the inability of this feedback to address the unexplored, individual learning needs:

"…you're not as engaged with it as if you were discussing it a bit more. I suppose it was just 'cause it was quite repetitive, I just kind of zoned out, when it was like the same things." (A30, female, 22)

Group B participants used many different adjectives that portrayed the learner in the active role. This group described their involvement in the learning process as 'probing', 'reflecting', 'analysing', 'processing', 'thinking', 'making changes'. In this model, the learners see themselves as the generators of ideas. When reflecting on how the feedback model affected their learning during the study, two different sentiments specifically relating to engagement of the learner are expressed. Firstly, they report that this model of feedback engaged them in a feedback discussion; they were required to contribute and participate in the feedback sessions, rather than to adopt a role of passive recipient. Secondly, they report a cognitive engagement with the task specifically as a result of the feedback model employed:
"So, I had to think about exactly what I'd done and then what it was that, in doing that, that made it go wrong or could make it better. So I had to think about it and I think, making me think about it, made me understand the problem better but then also meant that I was more likely to remember to do it next time." (B23, female, 22)

The process of cognitive engagement in the psychomotor task related to other positively-framed ideas associated with validation and realisation of competence and self-efficacy:
"Having my own voice in there was stronger than somebody else saying 'You know, you should move your hand this way or you should stretch the vessel that way'." (B19, male, 33)

Subsequently, participants in Group B articulated ideas relating to independence of learning and feedback. In this activated role, the learner moves beyond the enactor of tutor-driven or facilitated development and change, to the originator of independent reflection and progress:
"I think the techniques that you've been using, I could probably...if I actually gave myself proper feedback, sit down and go, like, if I got that piece of work and it was wrong I could sit down and say what went wrong or if I went into the history and felt it was like a bad history, I could sit and think, 'Why was it bad?' And try and pick out the bits like you were doing. So, like, play your role and play my role." (B23, female, 22)

However, Group B participants also described negative attributes associated with an active role in feedback (five participants, six items). This discussion centred around feelings of

frustration and a perception of pressure. Participants specifically related this to difficulty ‘remembering’ what had occurred during task completion and, therefore, may be associated with failures of intra-task monitoring and self-regulation. It highlights the importance of this stage of self-regulation in productive and acceptable dialogic feedback:
“So, I think the thing that I found the most tricky bit was trying to actually remember what I did in the task. I was sort of frustrated at myself because I couldn’t remember, ‘cause I was like, I’ve just done this.” (B14, female, 23)

9.2.1.2 Sources of feedback

Participants in Group A (information transfer feedback) were most likely to cite the tutor as the main source of feedback (n=24), whereas those in Group B (dialogic feedback) were more likely to cite either co-constructed tutor-learner discussion (n=13) or themselves (n=6) as the main source of feedback. These perceptions relating to source were consistent with the roles adopted during feedback, as previously discussed.

Participants in Group A commonly discussed the importance and value of objectivity in tutor-generated feedback. This draws an association between feedback and episodes of assessment, which may have been felt to be concurrent on occasion.
“Because when I was doing the task, I’m quite focused on what I’m doing. Obviously, I might be biased to what I feel I might be doing. So, someone who’s not doing the procedure assessing me at the same time giving me feedback will be more objective.” (A26, female, 23)

A significant portion of the participants in Group B identified themselves as the dominant source of feedback (n=6). This was not seen in Group A. This may relate to a learner perception of increasing independence, as most continued to make some reference to the role of the tutor:
“Towards the end of it, I thought the feedback came from me because I generated my goals, I thought about it at home, decided if I followed the goals in the session.” (B26, female, 24)

However, a significant subgroup of Group B participants (n=8) perceived the tutor as the main source of feedback during the study. Although some participants cited the tutor as the clear primary source of feedback, others in this subgroup did not completely negate the role of the learner but there was an apparent reluctance to detract from the role of the tutor:

"I think you were the main source of making me think about feedback. Sort of like, without you, I wouldn't have had any feedback but because you were like...you were one of the main sources of feedback but you made me think about it rather than...it wasn't just telling me feedback." (B23, female, 22)

9.2.2 Evidence of participant self-regulation

Group B (dialogic feedback) participants were more likely than those in Group A (information transfer feedback) to reference and provide examples of self-regulation of task performance through planning, intra-task monitoring and reflection. Conversely, Group A participants were more likely to discuss episodes where they displayed an inability to self-regulate task performance. However, a similar number of participants in each group displayed evidence of engaging in self-regulation outside of study visits, suggesting that this facet of self-regulation was less directly influenced by the feedback model (Table 83).

Table 83: Quantitative summary of density of self-regulation coding analysis

Evidence relating to self-regulation	**Group A** *Participant number (coded items)*	**Group B** *Participant number (coded items)*
Planning	4 (4)	16 (23)
Intra-task monitoring	12 (15)	21 (28)
Reflection	13 (23)	23 (46)
Out with study visits	6 (7)	6 (9)
Absence of self-	9 (17)	0 (0)

9.2.2.1 Planning

In relation to coding density, there was a disparity between the groups in relation to evidence of pre-task performance planning (Group B n=16; Group A n=4). However, interestingly, the participants who discussed specific incidences of planning, did so with a focus on process rather than outcome, regardless of group. This would suggest that Group A participants who tend towards process-driven task engagement are more likely to engage in the planning element of self-regulation, independent of feedback approach. This may explain why Group B participants were more likely to cite instances of self-regulatory planning; having engaged with a model of feedback focused on process. Furthermore, the discussion surrounding self-regulatory planning contained a high proportion of specific references to consideration of process goals, suggesting a link between construction of process goals and facilitation of pre-task performance planning.

"I think there was … how to have the vessel position. So I need to think very specifically through what it is I need to do myself and sort of making my mental list of that before I go on and do it." (A21, female, 30)

"And then as the weeks went on then I will be thinking about it before I do it. Even like on the bus in, I'm like "Ah, yeah. Do this." So it's doing it before and then doing the process for each of those little many goals like doing it and checking as you go along." (B11, female, 22)

9.2.2.2 Intra-task monitoring

Although there was a greater volume of items relating to intra-task monitoring identified in the analysis of Group B interviews, the number of participants engaging in specific intra-task monitoring in Group A was not insignificant. Participants from both groups gave very specific descriptions of instances of intra-task monitoring. However, analysis of Group B participant interviews showed evidence of application of process-driven thinking to problem solve, which was not seen in Group A participants:

"So then as you're doing it, you're thinking like, is that clip applicator in the right orientation? Is that we were talking about? Have I moved that vessel forward?" (A16, female, 23)

"Because then like you know 'oh… you've not put the clip on properly because you rushed and you didn't have the clip applier in the right angle'. So you change that. Rather than just keep forcing the clip on again and again." (B16, female, 21)

9.2.2.3 Reflection

The discussion of process related to three specific ideas: specific instances of reflection on the task, the interaction between feedback and reflection, and the importance of reflection.

Participants from both groups gave specific examples of process-focused and task-specific reflection:

"So it was actually like, for example, the clip jaws, they weren't perpendicular and then thinking how...the best way to do that. So I was going through what I was doing and then going through the better way to do it." (B21, female, 22)

However, Group A participants also tended to engage in outcome-focused reflection, during which they would consider the success or failure of performance markers (such as rupturing the vessel, or wasting clips). This focus was absent from Group B participant discussion of reflection.

Similarly, participants in both groups reported that the feedback given or constructed could be used to facilitate reflection:

"Because after the first feedback session with you, going through the form, it then kind of puts the things in your mind." (A16, female, 23)

However, for participants in Group B this facilitation of reflection was linked to a feed-forward effect and future planning; development of a cyclical process of self-regulation. This seems to particularly relate to the metacognitive engagement of those participants during reflection:

"It was targeted at 'Can we think about ways to improve that?' So it wasn't even you telling me what I had to do, it was you leading me to think 'Oh right, okay. Well, actually the tool works like that. Actually, maybe the vessel wasn't straight on to me. I should do it this way next time.' So it was...that was really good that you led me to figure out what was wrong and helped me understand how to fix it." (B12, female, 23)

Finally, Group B participant discussion illustrated instances of consideration of the value of reflection and related this to a positive attribute of the dialogic feedback model. These participants were aware that this model of feedback relied specifically on their engagement in reflection of performance, and perceived a powerful positive effect secondary to this engagement:

"I'm not patient for touchy-feely airy-fairy stuff but I feel like I actually was really good and in terms of, it made you stop and think. And by actually processing all the feedback." (B16, female, 21)

9.2.2.4 Absence of self-regulation

In this section of the thematic analysis, interviews were coded for discussion relating to participant difficulty in demonstrating self-regulatory behaviour. It is noteworthy that such illustrations were only observed in Group A. Analysis illustrated two important problems or barriers: a reliance on tutor direction and explanation, and inability to retain feedback information between study visits.

Some of the participants in Group A maintained the belief throughout the entire study that understanding of the task relied upon tutor explanation, and improvement upon tutor direction. This is in keeping with the previously described passive learner role dominant amongst Group A participants (section 9.2.1.1):
"You have to understand what is going on but in order for you to understand what is going on, you need someone to tell you." (A7, female, 23)

However, most items coded within this node relate to the difficulty Group A participants acknowledge with retaining feedback information between visits and, therefore, the limited ability to carry knowledge and improvement from a previous to subsequent visit. This concept is valuable and interesting when viewed in the context of the inter-visit quantitative results (Chapter 8, sections 8.2 and 8.3), which depict a deterioration in performance in Group A but not Group B when comparing the last task performance of a previous study visit with the first task performance of a subsequent study visit. This observation is at odds with the narrative given by a section of Group B participants, who specifically highlighted the value of cognitive engagement in the task allowing them to build a framework of understanding that served as a scaffold for maintenance of task performance (section 9.2.1.1). The difficulty was noted to be particularly palpable during the visit four task performance (following a four-week hiatus):
"Because the clips wouldn't go on properly and I said 'Oh, this totally happened before. I can't remember how to solve this.' Because we discussed it one of the other times. And I said I can't remember what the solution was. I knew that I couldn't keep on clipping the vessels because I knew that's what I'd done before. But I couldn't remember exactly what I was meant to do." (A16, female, 23)

9.2.3 Evidence of and ideas relating to the use of goals

Participants from both groups discussed goals. This content can be considered in two broad categories: ideas and perceptions regarding goals, and the reported use of goals within the study.

9.2.3.1 Ideas and perceptions regarding goals

Participants in Group A referenced the value of having ‘general’ but also ‘measurable’ goals related to outcome, which helped define task completion. When considering the focus of these goals, there was evidence that these participants felt there was a motivating value to assessing the success of subtask outcome goals (applying a clip, not breaking the vessel), in addition to the over-arching outcome goal of task completion:

“I wanted to get one clip on each end and cut it. I didn’t want to be faffing around. So then I think I was probably more focused than had I just been like ‘oh well, I’ve just got to cut the vessel’.” (A16, female, 23)

Interestingly, there was discussion of the potential negative effects of using outcome goals, although this discussion was vague and it appeared that the limitations of using over-arching outcome goals did not translate to appreciation of similar limitations in using subtask outcome goals:

Participant: “For example, a goal of ‘I want to do well in the lap simulator’ is not very beneficial.”

Tutor: “No.”

Participant: “But when you have specific clear goals and aims, you’re able to target those.”

Tutor: “Give me an example of a useful goal you might have in relation to one of those tasks.”

Participant: “Right. So one of the areas I have an issue is with pace.”

Tutor: “Okay.”

Participant: “And so one of the goals was to kind of speed it up.” (A24, male, 22)

Participants in Group A noted the potential ‘demoralising’ or ‘frustrating’ effects of failing to achieve outcome goals:

"So I would say I want to… fully complete task number one, whether or not the stats are good but I want to complete it with no blood loss and it might not be realistic. It might result in frustration. It might result in missing out on some of the finer details, things that you could improve upon." (A12, male, 22)

In relation to process goals, which were less often discussed by Group A participants, learners reflected that achievement in relation to outcome was linked to adopting a more instruction-based approach to the task:
"Well for me in this task, an achievable goal is getting the clip on first time on my first task and I found that difficult throughout. But I proved to myself that it was achievable with manipulation of my skills. So I had to focus on certain aspects of what I was doing, the way I was positioning the vessel, and rotating my clip and that was probably the main thing I focused on because that was the main thing that was holding me back in the task." (A12, male, 22)

In contrast, participants in Group B were more likely to express ideas in relation to process than outcome goals (19 vs 12 coded items). When discussing outcome goals in particular, participants in Group B focused exclusively on the negative effects of employing these goals and the negative attributes that limited their application to the task. Outcome goals were described as 'too broad', 'inhibiting' and 'not instructional', with learners reluctant to relate them to improved task performance:
"Because the whole point is you might be struggling to get the clip on properly, having the goal of 'I'll just get the clip on properly' doesn't actually help at all." (B30, male, 22)

Conversely, when discussing ideas relating to process goals, participants in Group B were more likely to use language that described their positive attributes: 'specific', 'constructive', 'instructional' and 'well-defined'. When specifically comparing outcome and process goals, participants promoted the use of process goals, concluding that they were more readily utilised when performing the task. Group B participants drew links between the use of process goals and reflection, the building of confidence and increased focus during task performance:
"[My goals] were descriptive enough that I understood exactly what I had to do and what I have to put in practice in the future particularly in between sessions rather than just from task to task." (B24, female, 27)

9.2.3.2 The reported use of goals

Group A participants predominantly employed outcome goals during the study (23 participants, 41 coded items). These outcome goals tended to focus on 'successful' completion of specific subtasks (not breaking the vessel, using the minimum number of clips) but also included general attributes that defined a good performance (minimising damage to the abdomen, moving quickly).

"I tried to get a clip applied within a sort of set number of tries. First clip is what I went for but I said to myself I want it under three." (A29, male, 24)

However, there was also evidence of Group A participants employing process goals in relation to the task performance. These participants described the process of breaking down the task into smaller, instructional components, which helped them to complete the task accurately. Interestingly, this content differed from Group B discussion relating to the use of process goals in two ways: it was often reported to be tutor-generated and Group A participants did not necessarily identify process goals as goals:

"I was really struggling because I couldn't see both side of the clipper and you said 'Make sure you put this all the way into the hilt and then if you turn it slightly… you will be able to see both sides of it'. And then when I did it the next time it was obvious that that was better." (A21, female, 30)

There was extensive discussion relating to the reported use of process goals in the Group B participant interviews (21 participants, 27 coded items), with no reported use of outcome goals. This would suggest that whilst the dialogic feedback model likely encourages the use of process goals, it almost certainly discourages the use of outcome goals. When discussing their use of process goals, Group B participants used particular descriptors: 'specific to individual', 'specific to task', 'prompts' and 'instructions'. Group B participants were able to articulate clear examples of process goals used within the study and how they were facilitated by cognitive engagement in the task:

"Okay, for example every time I went to cut the vessel itself was moving about quite a lot. And I changed that by setting myself some very simple goals, to hold the vessel still in the position I wanted it before even attempting to cut it, then face the scissors to me and ensure that I was using the middle of the scissors as oppose to hilt. And just doing those two things changed the outcome completely." (B26, female, 24)

Group B participants related the successful completion of process goals with completion of outcome goals:

"So, I think I often kept the same goal like after my second visit when I was to orientate the devices so that I can see them in front of me. And that's really a simple thing to do but I thought that really helped me throughout because it just helped me just remind right at the start to do that and then as soon as you've done that, it's much easier to, I suppose, complete the task." (B30, male, 22)

Participants in Group B commonly reported the use of process goals to create very prescriptive sequencing for the task:

"I think a useful aim from the tasks has been specific and it's spelled out exactly all of the steps and in glorious technicolour. So, going into the minutiae of check the position of the clip appliers before you do anything else. And then, step two, start with your left hand rather than your right hand to speed things up. And then step three, make sure you don't stretch the vessel. It's specific and it's small." (B19, male, 33)

They also discussed the link between development of process goals and a very specific, articulated intra-task monitoring process. In some examples, the *completion of the process goals is being checked*; in others, the *checking process itself becomes the subject of a goal*:

"But then the other one sort of like I think I made a goal about pulling the vessel in a certain way. Well I always thought about [vessel movement] being upward and towards me. Like that was a good goal because it was specific and like I can check that." (B27, female, 21)

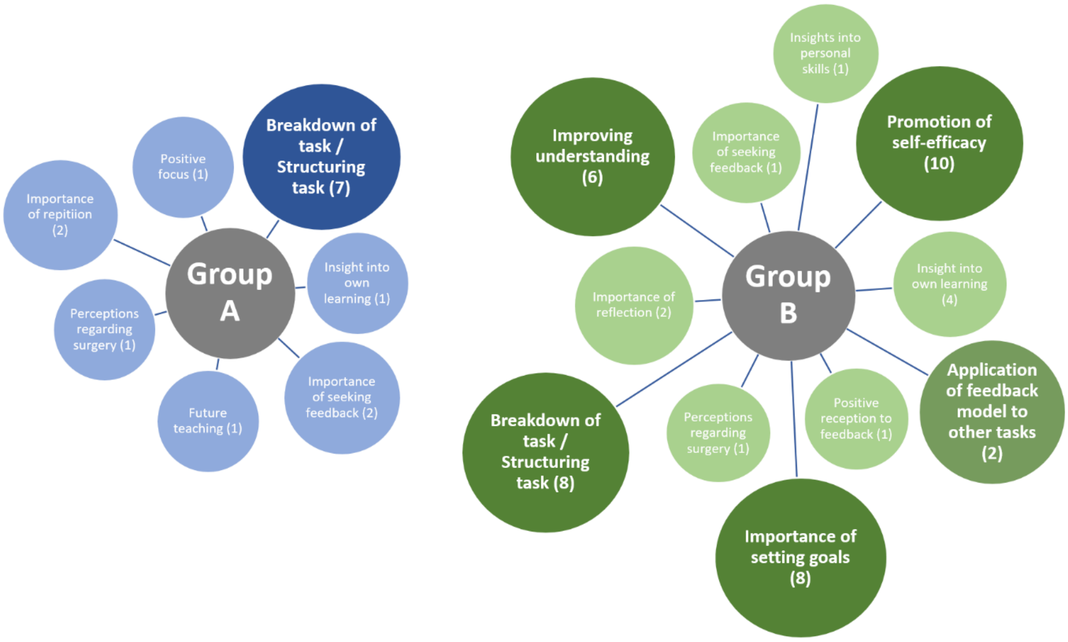
9.2.4 Changes in participant perception and ideas relating to future learning

Thematic analysis explored participant ideas in relation to changes in participant perception regarding feedback. The ideas identified, and the frequency of their representation, is illustrated by feedback group (A and B) in Figure 64. A similar number of participants in each group expressed ideas relating to a change in perception of feedback as a consequence of the study (Group A n=13; Group B n=15). However, Figure 64 illustrates that participants in Group B (dialogic feedback) expressed a greater number of different ideas (11 vs 7).

Participants in Group A (information transfer feedback) were most likely to describe changes in their perceptions relating to the usefulness of breaking down and structuring the task following participation in the study. This may relate to the highly structured nature of the information transfer feedback. Only small numbers of participants in Group A expressed ideas in relation to other changes in perception.

Participants in Group B (dialogic feedback) expressed ideas in relation to improving understanding, the importance of the promotion of self-efficacy, the usefulness of breaking down and structuring the task, and the importance of setting goals following participation in the study. A smaller number of participants expressed ideas in relation to other changes in perception.

Figure 64: Exploration of themes in changes in perception (with frequency)



9.2.4.1 Breakdown of the task & Promotion of task structure

Participants from both groups reflected that the feedback in the study had increased their awareness that breaking down a psychomotor task into subtasks and creating structure would improve their task performance and that this was a practice they could apply to future learning situations. It would appear that both highly-structured information transfer and dialogic feedback could invoke this perception:

"And it's just given me a new way of thinking about these tasks that I'm going to do and there are many more that I'm going to learn. It taught me how much I do value the breaking down element." (A10, male, 25)
"I think the importance of like really breaking them down especially like a new task that you're not too confident into really small, sort of like chunks or something." (B14, female, 23)

9.2.4.2 Promotion of self-efficacy

A new appreciation of the importance of realising self-efficacy was a strong theme discovered amongst Group B participants. It was not reported by Group A participants. Group B participants reflected that the feedback model used forced them to 'come up with the answer' by 'thinking it through' themselves. This strongly echoes the evidence relating to the active role often adopted by Group B participants (section 9.2.1.1) but, beyond this, it suggests that the dialogic feedback model promoted awareness of self-efficacy, and participants valued this consequence:
"Actually, I probably knew the answer so really I should ask myself if I can improve things actually and that I probably do have the answer in there if I look at it and analyse gradually." (B11, female, 22)

9.2.4.3 Promotion of task understanding

Another of the ideas expressed strongly and exclusively by Group B participants was that the dialogic feedback model encouraged a deeper understanding of the task itself and that the active role of the participant during feedback relates to task understanding, development of self-efficacy and subsequent independence of reflection (section 9.2.1.3).
"I think the main strength of the feedback that I've had during the course… was to kind of put it in terms of context. So, understanding the practicality of each of the stages of the tasks and being able to see why certain choices would lead to different consequences and why that it had caused problems in previous attempts..." (B24, female, 27)

9.2.4.4 The importance of goal setting

Emphasis on the creation of goals in relation to task reflection, planning and, ultimately, performance was a feature of the dialogic but not the information transfer feedback model. It is, therefore, not surprising that reflections relating to the importance of goal setting was observed in Group B but not Group A participants. However, that it was such a strong sub-theme relating to changes in perceptions indicates two ideas: that previously these participants did *not* highly value goal setting; and that goal setting was perceived as a useful element of the dialogic feedback model:

"So originally on my first session, my aims and goals were to pretty much just be the best and to complete the task. But that's not an actual goal or aim that will ever help me to achieve that. So it's important to have aims and goals but it's important to have ones that are helpful rather than putting pressure on yourself to achieve a final outcome." (B16, female, 21)

9.2.4.5 Wider application & Feedback literacy

A sub-theme less frequently discussed but noteworthy in relation to its implication for participant learning, was application of the dialogic feedback model in psychomotor task performance out with the study. Two of the Group B participants related narratives of employing self-regulatory skills to tasks encountered during their medical studies. This was not observed in Group A participants, suggesting it was a phenomenon particular to the dialogic feedback model and not an effect produced by the information transfer feedback model:

"But I think the thing that helped me with the feedback is like thinking about well it's a learning process… And so I think I've sort of done that a bit with sort of my approach to cannulation already… And then when I don't succeed, think about how I can make it better. So there was this one point where I forgot to put my finger over the vein you know to keep blood from pouring everywhere. So even though I got the cannula in, it was a bit messy. And so instead of thinking about that I completely failed you know instead I thought 'Okay, well next time I just have to make sure that I am conscious of doing that step'. I think what this feedback has really helped me in my thinking about ways to approach my learning." (B2, female, 28)

9.3 SUMMARY

The qualitative element of this study has explored participants' experiences of feedback and highlighted their conceptual understandings and learnings. Participants exposed to the dialogic feedback model cited a co-constructed tutor-learner dialogue as the dominant source of feedback during the study and were more likely to adopt an active role during the feedback model than those in the information transfer group, who felt that the tutor was the main and most important source of feedback. The dialogic feedback group described a learning process in which greater understanding and cognitive engagement in the task and realisation of self-efficacy could lead to independence of learning.

Using the dialogic feedback model enabled participants to engage in the self-regulation cycle of reflection, planning and intra-task monitoring. These behaviours were reported by participants exposed to the information feedback model to a lesser degree, as were behaviours relating to failure of self-regulation. The information transfer model seemed to promote students' use of both outcome and process goals but the former was more prevalent. The dialogic feedback model was associated with use of clearly articulated process goals and ideas relating to their value and contrasting ideas of the limitations of outcome goals.

Furthermore, whilst both feedback models impressed upon the participants the value of creating structure within task completion, a more sophisticated understanding of feedback, the promotion of self-efficacy and the conscious use of process goals were valuable sequalae particular to the dialogic model of feedback.

CHAPTER 10: DISCUSSION AND CONCLUSIONS

Feedback is widely accepted as an important potential contributor to the process of learning (Carless, 2006; Veloski et al, 2006). Combined meta-analyses have identified over 100 different factors relating to feedback that affect student educational achievement (Hattie & Timperley, 2007). Feedback is an often-discussed subject within medical education where teachers feel pressure to provide feedback and learners expect to receive it.

The overall quality of the quantitative experimental studies is poor, limiting what this body of the literature can tell us about the effect of information-transmission feedback on learner performance. Within these studies, there is no exploration of the learner role and no studies involved a dialogic feedback process.

Within wider education a reconceptualisation of the feedback process has been considered, with a move from a one-way information transmission of information to the co-creation of feedback dialogue between tutor and learner. In this model, feedback is ‘exploratory’ rather than ‘directive’, the learner role is enhanced, and the onus moves away from the teacher’s role in delivering feedback, towards the learner’s active role in co-forming and internalising feedback to enhance understanding and performance. However, no link between an active role and learner self-regulation have been drawn.

The highly structured and comprehensive feedback literature review presented in this thesis illustrates that an advanced, dialogic model of feedback is currently only in the concept stages within medical and surgical education. The surgical education literature remains preoccupied with exploration of the effect of tutor-generated and delivered feedback via experimentally designed studies. The overall quality of these quantitatively-focused studies is poor, limiting what this body of literature can tell us about the effect of information-transfer feedback on learner performance. Within these studies, there is no exploration of the learner experience or role and no studies involved a dialogic feedback process. Discussion within the medical education literature of new ideas and concepts have theorised a paradigm shift towards a shared tutor:learner feedback dialogue (Carless et al, 2011; Boud & Molloy, 2013) are limited by the tutor-centric language that persists, and the lack of coherent understanding of how this feedback is created (Rudland et al, 2013; Telio et al, 2015).

Therefore, this MD project set out with the intent to investigate the effect of dialogic feedback and promoted self-regulation of learning of psychomotor task performance, skill retention and the learner experience of feedback. Two specific research questions were asked:

What is the effect of an integrated model of dialogic feedback with encouraged self-regulation versus an information transfer model of feedback on psychomotor task performance and longevity of skill retention?

What is the effect of the integrated model of dialogic feedback on learner's experience and understanding of feedback?

10.1 SUMMARY OF KEY FINDINGS

10.1.1 Research question one: The *What*

What is the effect of an integrated model of dialogic feedback with encouraged self-regulation versus an information transfer model of feedback on psychomotor task performance and longevity of skill retention?

Both information-transfer and dialogic feedback models were associated with improved psychomotor task performance, in relation to both efficiency and accuracy. In relation an information-transfer model of feedback, these results support other work in the literature that suggested more efficient and accurate psychomotor task performance following individualised instructor summary feedback when measured using objective measures (Paschold et al, 2014). Conversely, the findings contradict other work in the literature that suggested that this model of feedback was not associated with improved performance of psychomotor tasks when quantified by objective measures (Rogers et al, 2000; Kruglikova et al, 2010, Boyle et al, 2011; Farjad et al, 2013).

There is no existing evidence in the literature that has studied the quantitative effect of a dialogic model of feedback. Therefore, the evidence that is presented here suggesting that it too is associated with more accuracy and efficiency, serves to extend the body of evidence.

Additionally, the randomised control trial design allows further extension of the literature via direct quantitative comparison of the effect of the two models of feedback.

The results of this study suggest that participants who engaged in a dialogic model of feedback performed the task more efficiently at the start of study visits two, three and four, immediately following shorter (one week) and longer (four week) breaks from task performance, than those engaged with an information-transfer model. The association between information-transfer feedback and deterioration in performance at delayed testing contradicts the finding of Porte et al (2007); who reported maintenance of performance at delayed testing associated with summary expert feedback.

Over the course of the study, participants who engaged in a dialogic model of feedback demonstrated a significantly reduced intra-group variation in performance as compared to those engaged in an information-transmission model. This is previously unreported in the literature.

None of the studies in the existing literature feature a cross-over psychomotor task. A cross-over task is one related to the primary study task but one in which the participant does not receive direct feedback. The purpose of including a cross-over task is that inter-group comparison of performance conveys information about a participant's ability to apply understanding in a wider context, and offers insight into learning. The results of this study suggest that learners engaged in a dialogic model of feedback perform related tasks both more efficiently and accurately than those engaged in information-transfer feedback. This would suggest that a dialogic model of feedback promotes a learner's ability to apply understanding out with its initial and primary context but does not suggest the same association is evident with information-transfer feedback. This association has not previously been investigated or reported in the literature.

10.1.2 Research question two: The *Why*

What is the effect of the integrated model of dialogic feedback on learner's experience and understanding of feedback?

10.1.2.1 Learner roles & sources of feedback

This study suggests that learners engaged in a dialogic model of feedback are more likely to cite themselves as the primary source of feedback, or report co-created feedback with a tutor, than those engaged in an information-transfer model of feedback, whom are more likely to report that the tutor as the primary source of feedback.

This recognition of the tutor as the primary source of feedback may relate to learners engaged in an information-transfer model of feedback being more likely to adopt a passive role during feedback. In contrast, learners engaged in a dialogic model of feedback were more likely to adopt an active role, associated with evidence of greater cognitive engagement in the task, realisation of self-efficacy and a recognition of themselves as the originator of independent reflection and practice.

The active learner associated with the dialogic feedback model bears resemblance to the portrait painted by Riordan and Loacker (2009), as an '*independent lifelong learners who have learned from us but no longer depend on us to learn'* (p. 181). Exploration and application of the dialogic feedback model serves to extend our understanding of a less 'narrow and transmissive' view of learning (Ajjawi, 2012) and appreciate in practical terms how the co-constructed model of feedback championed by Telio et al (2015) might be fostered in real life practice.

This qualitative analysis relating to learner roles may offer some triangulation and explanation of the quantitative results of this study. The greater cognitive engagement of the learner in the task fostered by the dialogic feedback model may explain both the increased consistency and maintenance of improvement between study visits observed in these learners.

10.1.2.2 Evidence of participant self-regulation

This study suggests that involvement in the dialogic model of feedback resulted in learners being more likely to give examples of self-regulatory behaviours than those involved in an information-transfer model of feedback. Conversely, the latter were more likely to give specific examples of having difficulty in adopting self-regulatory practice.

Whilst learners engaged in either dialogic or information-transfer models of feedback described the intra-task monitoring facet of self-regulation in similar volumes, more evidence of the planning stage of self-regulation was seen in the learners involved in dialogic feedback. Learners involved in information-transfer feedback were more likely to cite difficulties remembering tutor directions or retaining feedback between study visits. The dialogic feedback learners were also more likely to describe examples of reflection that focused on process, with emphasis on specific ways to improve future task performance and problem solving. Whereas learners engaged in information-transfer feedback were more likely to describe reflection that focussed on task outcome, rather than process.

This evidence, suggestive of greater engagement in planning and process-driven reflection, may explain both the improved inter-visit maintenance of performance and decreased variation in performance observed in the dialogic feedback learners.

These findings both support and extend the current medical education literature. Sobral (2005) and Artino et al (2014) both reported improved academic performance in relation to increased self-regulatory activity. The results of this study would suggest that the same is true in relation to psychomotor tasks. The results of Brydges et al (2012) suggested that psychomotor task skill retention was improved by directed self-regulated learning rather than instructor-regulated learning. This study supports this finding (more efficient and accurate task performance at delayed testing) and may help articulate why this is the case.

10.1.2.3 Evidence of and ideas relating to the use of goals

No previously published work has featured the application of a dialogic model of feedback with encouraged use of process goals, although the formation of goals and even the emphasis on process goals, is part of several suggested models of self-regulation (Zimmerman and Moylan, 2009; Sandars, 2011; Brydges & Butler, 2012). Therefore, analysis of the qualitative results of this study contributes new information to the body of literature.

Learners engaged in an information-transfer model of feedback were more likely to describe outcome goals when asked to give examples of goals relating to performance of the study task. However, they were also likely to use emotional descriptors with negative connotations when describing the potential use of this type of goal, such as 'demoralising' and 'frustrating'. In contrast, learners engaged in a dialogic model of feedback were more likely to give examples of process goals in relation to task performance. Additionally, the dialogic learners were more likely to describe avoidance of using outcome goals, and use analytical descriptors in relation to them, such as being 'too broad' or 'non-instructional'. These findings suggest that a dialogic model of feedback may be an effective method in promoting the use of process-related goals and that learners engaged in this feedback model may develop greater learning literacy in relation to the use of process and outcome goals in psychomotor task performance.

10.1.2.4 Changes in learner perception

Both groups of learners, engaged in information-transfer and dialogic feedback models, described changes in perceptions relating to learning. In particular, both groups of learners described a newly-appreciated value in feedback promoting a breakdown of the task and a greater understanding of a structured approach to task completion. However, those engaged in dialogic feedback displayed a greater volume of and variation in their responses.

Dialogic learners frequently described changes in perception in relation to the importance of goal setting, of understanding the task (not just appreciating its structure) and realisation of self-efficacy. These changes in perception were not seen in the information-transfer learners. This study, therefore, may suggest that engagement in a dialogic feedback model provides a greater challenge to the learners' ideas of feedback, increases feedback literacy and promotes a reconceptualisation of feedback within the learner.

10.2 A PROPOSED DIALOGIC LEARNER MODEL OF FEEDBACK

Analysis of interview discussions helped explore the active learner role and evidence of self-regulation, and their relationship with how the feedback was utilised. Figure 65 illustrates a proposed (but most likely immature) dialogic learner model that was generated via analysis of the ideas expressed by participants when adopting an active role during feedback and engaging in specific self-regulatory behaviours relating to task completion.

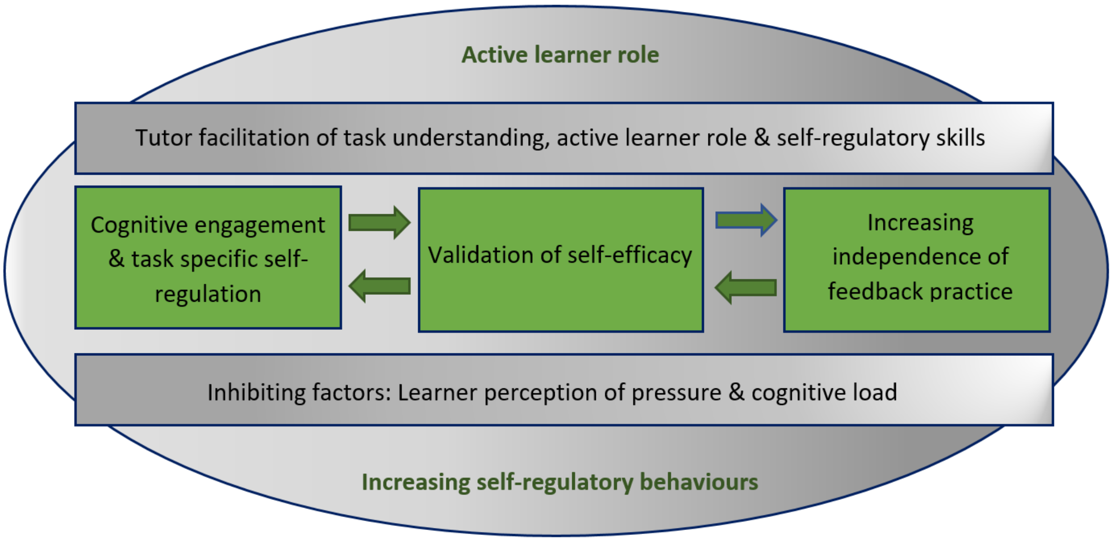
The reflections of the participants suggested that there was an interaction between creation of an active learner role and cognitive engagement with the task, increasing perceptions of self-efficacy and, later, independence of feedback. The difference between the learner observations that typified the information-transfer and dialogic feedback groups suggested that the role played by the tutor during feedback regulates activation of the learner role.

The dialogic feedback model enabled better participant engagement in the self-regulation cycle of reflection, planning and intra-task monitoring. It was associated with use of clearly articulated process goals and this is incorporated into the model.

A key feature of this model is that it is learner-centric. The tutor is described in a faciliatory role. Their role is specifically to facilitate the learners' understanding of the task, their active role during feedback, and development of self-regulatory skills. The tutor is not to contribute value judgements regarding observed task performances. The potential effect of learners struggling with the cognitive load of self-regulation is recognised as a potential barrier to adopting an active learner role during feedback but this reduces with practice and increasing perception of self-efficacy.

In the first stage of the model, emphasis is placed on learners' cognitive engagement in the task with explicit encouragement of task specific self-regulatory behaviours (planning, intra-task monitoring, reflection with focus on formation of process-driven goals). Via increasing learner understanding of the task achieved via persistent cognitive engagement and repeated learner-driven self-regulation, learner feelings of self-efficacy are promoted. In the final stage of the model, the faciliatory role of the tutor is diminished as the ability of the learner to engage in productive self-regulation becomes independent.

Figure 65: Proposed dialogic learner model of feedback



The proposed dialogic learner model of feedback provides a fully articulated and practical model of dialogic feedback. This is new to not only the medical educational literature, but also to the wider educational literature. However, further study in relation to psychomotor task learning would be required to enhance understanding and allow refinement of the model. Further qualitative research focusing on the learners' experience may be most helpful in this next step.

Furthermore, it must be acknowledged that the proposed model is based upon observation and analysis of a dialogic feedback model when used in psychomotor task learning in a simulated environment. Caution would be advisable when considering its application to different learning activities, such as written assignments; or different learning environments, in which learning was less structured and more episodic, or when more than one tutor is involved in the learning process.

The proposed model differs from Sargeant's model of self-directed assessment (Sargeant et al, 2008) in that feedback is co-created via learner:tutor discussion. The role of 'facilitation' is attributed to the tutor, is expanded and better articulated. The 'use' of feedback mentioned in step three is replaced with ongoing self-regulation.

The model (Figure 65) is student-centred, as is the one previously proposed by Rudland et al (2013) but removes the necessity for the student and tutor from the ‘recipient’ and ‘provider’ roles in the feedback process. Rather than focusing on the desirable qualities of these two agents, it focuses on the results of their interactions.

The proposed dialogic learner model of feedback supports Carless et al’s (2011) view of a reconceptualised dialogic feedback: the learner role is enhanced, feedback discussions are exploratory rather than directive, and they are used to create a scaffold from which learner independence can be achieved.

10.3 STRENGTHS

10.3.1 Design & generalisability

The quantitative element of this study utilised a randomised control trial (RCT) parallel-group design. Within clinical medicine, RCTs are considered to be the most powerful experimental design. The randomisation process serves to reduce sources of bias that may be encountered via other designs, such as cohort or case studies. When the two randomised study groups were analysed, there were no significant inter-group differences in relation to the pre-study experience, pre-study self- or tutor-confidence, hand dominance or age. There were no significant inter-group differences in relation to post-study participant self-confidence, confidence in the tutor or satisfaction with study participation, eliminated these as possible sources of bias.

The rigour and quality in relation to the design of the study were promoted by conducting extensive pilot studies. The pilot studies allowed for trial and adaptation of study materials, such as participant task information sheets, presentation and videos. They also allowed opportunity for development of the feedback intervention utilised in each study group, with maturation of the technique of employing each feedback model at each reiteration.

Adherence to study protocol in relation to timing of study visits was accurate and inter-visit intervals were similar between groups. Similarly, there were no significant inter-group differences in relation to feedback volume.

All feedback episodes and data collection was performed by the lead researcher, which reduced variation in tutor performance, promoted adherence to study protocol, and the integrity of feedback model presented during feedback sessions.

Participants were judged to have completed the study if they completed the first three study visits. Participant retention was excellent during the study, with only seven of the starting 68 failing to attend one of the first three study visits. This equates to a participant completion rate of 90%. Of the 61 participants included in the final study cohort, 56 (92%) completed the delayed retention testing and semi-structured interview in study visit four. The high participant completion rate promotes the quality of the study and the validity of its results and conclusions.

10.3.2 Validity & reliability

Significant efforts were made to promote validity of the quantitative study results. Multiple measures of performance, both in relation to accuracy and efficacy, were considered when selecting the psychomotor task to be used and a method that allowed automatic, objective computerised-calculated measures was chosen. Furthermore, the measures of performance used were congruent with how surgical performance might be judged in clinical practice, with economy of movement, time efficiency and avoidance of errors all signs of skilled operating.

10.3.3 Triangulation

Triangulation of results and conclusions was promoted by three design features: the repeated collection of multiple measures of quantitative performance; the inclusion of a cross-over task; and a mixed methodology allowing deductive, quantitative data to be interpreted alongside inductive, qualitative data. In combining both quantitative and qualitative design elements, this study represents the only mixed methods study within medical education in relation to the exploration and quantification of the effects of a feedback model.

10.4 LIMITATIONS

10.4.1 Quantitative study elements

10.4.1.2 Generalisability

Although the randomisation of participants to each study arm allowed for the controlling of most participant demographics, there was an inequality in relation to gender. There was significant difference in the proportion of male and female participants allocated to the groups ($p=0.049$), with a greater number of female participants allocated to the dialogic feedback group.

The voluntary recruitment of undergraduate medical students from a single university may represent confounding factors in relation to study design and limit the conclusions that might be drawn from the results. It is possible that the students recruited to the study, which involved a not insignificant time commitment, may represent the more motivated undergraduate and learners and, therefore, it cannot be guaranteed that the convenience sample obtained here is representative of undergraduate medical students as a wider group. It is not possible to know what effect this may have on the results of the study.

Furthermore, the results of this study and the subsequent conclusions drawn relate only to the naïve learner. None of the participants were experienced in laparoscopic surgery (either real or simulated) prior to the start of the study. Therefore, the results relate to the early stages of learning a psychomotor task. It is not possible to say whether the associations between feedback model, self-regulation and task performance seen in this study would be replicated in more advanced learners.

Due to the experimental design of the study, task performances and feedback sessions were conducted within a controlled environment. Whether the results found here in relation to feedback model and psychomotor task performance would be replicated with their introduction into educational practice is unknown.

Finally, the position of the lead researcher as a surgical trainee and undergraduate teacher cannot be quantified. These positions may influence participant behaviour. This may promote engagement with the study, the feedback sessions and participant motivation. Whilst these may be viewed as ‘positive’ effects, they represent a potentially confounding effect when considering generalisability of the study findings to wider practice. However, such an effect would have applied to both groups equally.

10.4.2 Qualitative study elements

10.4.2.1 Credibility & conformability

<u>Whilst the semi-structured study design attempts to promote the voice of the participant, thereby promoting credibility, it must be acknowledged that the questions asked, and the interpretation of answers given, is presented here through the lens of the insider researcher. Whilst there was no attempt to deliberately misdirect interviews, nor interpret the data with overt bias, it must be acknowledged that the phenomenon of positive publication bias may result in researchers being more motivated to find positive or significant results rather than negative results (Murad et al, 2018).</u>

Similarly, whilst this researcher attempted to promote transparent and good research practice by avoiding uneven interview techniques between study groups, the risk of bias remains, such as use of different tone with participants from different groups, inadvertent body language, and the use of positive or negative language.

Whilst standardisation of questions between study groups was helpful in this study, in which there was no pre-existing framework for interpretation of results; it might also have served to limit discussion and future studies investigating the learner experience in relation to feedback model might choose an unstructured interview design to further promote conformability. These limitations might be mitigated by employing a design that included interviews conducted by an individual not involved in the study. However, the practicalities of the time required in this study and the difficulty of a person out with the study being able to engender the same quality of discussion and richness data are barriers to this solution. As an external check of rigour, the audio recording of five post study interviews from each study group are available for review and critique via the data disc included with this study. (Interviews randomly selected via Excels ‘random number’ generator function; see file ‘Full study results’, sheet ‘Participant interview review’.)

10.4.2.2 Dependability

To a degree, the truth of the qualitative data and its interpretation is protected by the triangulation afforded by the quantitative data. The ideas expressed by participants was viewed in the context of and related to these objective behaviours, so conclusions from induced, qualitative data were created and framed within the results of deducted, objective data, improving dependability.

However, no study is free of the inherent bias of the researcher, and choices regarding design, data collection and data interpretation ultimately will shape results. I would conclude that having acknowledged this inherent bias, the overall quality of interview conduction does allow a degree of confidence in the dependability of the data.

10.4.2.3 Transferability

When examining the transferability of the data and interpretations, acknowledging the participant population sample and position and reflexivity of the researcher is key. It is difficult to be sure that different a participant sample (for example, doctors at a different stage of training, or completely unknowing of the researcher) or different researcher (of different experience, training background, or perceptions) would yield the same data or conclusions, even if the same questions were asked. When considering that a different study design (for example, unstructured interviews or focus groups) could also have been chosen, the possibility for variation in data generated and conclusions drawn is almost infinite. Therefore, the fairest and safest interpretation of this data is to say that these findings were true in this study but should be corroborated or refuted but certainly augmented by further study in other settings and populations.

10.5 IMPLICATIONS

10.5.1 Future feedback practice within medical education

This study explores the practical aspects of creating dialogic feedback, the quantitative effects of this model, and its relationship with the learner experience of feedback. It demonstrates that the principles of dialogic feedback can be distilled into feedback characteristics (section 3.7, Box 4) and that these can be used to create exploratory, co-created feedback conversations and feedback. The researchers involved in this study hope that it provides the beginnings of a map that might allow for the introduction of dialogic feedback to medical education practice.

In relation to the quantitative effects of dialogic feedback on psychomotor task performance, the findings of this study suggest that it is a viable alternative to information-transfer feedback. Indeed, the results of this study suggest that dialogic feedback may confer some advantages in relation to both the efficiency and accuracy of performance, particularly over a prolonged period of learning, and when learners are required to relate skills acquired in relation to one task to other similar tasks.

Exploration of the learner experience of feedback suggests that a dialogic model promotes a more active learner role with associated promotion of self-efficacy and changes in ideas and perception of feedback. These factors may be of interest to those involved in both learning design and the practical delivery of psychomotor task teaching. Adopting such a model would require faculty development and a shift in the understanding of the purpose of feedback towards developing self-regulation rather than correction, as well as promoting a process-oriented dialogic conception of feedback.

10.5.2 Future feedback research

This work represents the first statistically robust study to explore the effects of information-transfer and dialogic models but the limitations of this study give rise to unanswered questions, which necessitate the need for further exploration.

Further study of the effects of a dialogic feedback model in more mature learners would be of interest, particularly in relation to post-graduate training. Studies of clinical trainees may be combined with a study based in everyday clinical practice, rather than the controlled environment utilised here.

Similarly, whilst the multiple study visits allowed each feedback model to be firmly established, the effects of dialogic feedback of a more episodic nature with less intense tutor input are unknown and may be highly relevant to medical education.

10.5.3 Dissemination & utilisation

As reflected by Tavakol & Sandars (2014), study results must be shared with the wider medical education community, in the hope that their use might improve medical education and, ultimately, patient care. I suggest that there are four levels at which this work may be disseminated to medical educators, each with their own benefits but also practical issues and barriers to utilisation.

The first is via journal publication and at this level, it is hoped that this work might contribute to the academic discourse surrounding feedback and self-regulation in medical education, and even the wider field of education. In doing so, this work would gain the benefit of peer review and critique and possibly spark further study of the ideas raised by this work. The practical difficulty in disseminating the work in this way is one of condensing such a large body of work into meaningful but digestible publications. Ones which neither lose their context by over-simplification, nor are incomprehensible due to length or required level of pre-existing knowledge on part of the would-be reader. Finally, I would assert that this level of dissemination has significant limitations in utilisation by the average medical educator, who is a practicing clinician, who does not regularly read medical education journals. I think considering this is also of importance when formulating articles for publication and considering target journals. Some texts may wish to appeal to the theoretical appetites of medical education academics, focusing on the development of the dialogic feedback model and discussion of evolving learner perceptions and feedback literacy. However, other papers may be more usefully targeted towards practicing clinicians and surgeons, who are more likely to be interested in the quantitative results of feedback, the practicalities of how different feedback models are engendered.

The second level of dissemination is that of presentation at medical education conferences. The likely target audience at this level is academic medical educationalists (likely with an established interest in feedback or self-regulation) but also more novice, would-be medical educators. It is this researcher's reflection that many junior clinicians with a developing interest in medical education may attend such meetings and this work may be more effectively disseminated to this audience via this strategy than through peer reviewed journal publication. The practical difficulties encountered at this level is the time and resources required to attend such meetings (by presenting researchers and audiences) and, again, the difficulty of distilling complex work into succinct ten-minute presentations. It would be necessary to split this body of work into more focused episodes, each with a clear method (for validity) and message (for utilisation).

The third and fourth levels of dissemination are speaking at local educational meetings and courses, and assimilating dialogic feedback practices into everyday practice and promoting a local discourse with other clinicians and educators on this topic. Whilst such meetings are likely to attract a self-selecting group of well-motivated clinical educators, attendance at educational courses is becoming increasingly required for both doctors in training and practicing clinicians and surgeons in demonstrating continuing professional development. Furthermore, clinicians regularly engage in both formal and informal learner feedback in day-to-day educational activities and 'water cooler' discussions might be valuable in garnering a change in practice, if educators can be convinced that this might promote better performance and patient care. This level of dissemination will not result in academic critique nor further theoretical development of the ideas presented but may be powerful on a local scale in relation to utilisation.

10.6 CONCLUSIONS

This study illustrates that it is possible to utilise a dialogic model of feedback when teaching a psychomotor task. This study presents strong quantitative evidence that such a model of feedback has beneficial results in relation to efficiency and accuracy of task performance when promoted over time. Therefore, a dialogic model of feedback represents a credible alternative to the information-transfer model currently employed in medical education.

The qualitative exploration of the participant experience of co-constructed dialogic feedback may suggest promotion of an active learner role and cognitive engagement, with increasing perceptions of self-efficacy and, later, independence of feedback. Contrasting learner observations between the information-transfer and dialogic models suggests an association between the role played by the tutor during feedback and activation of the learner role, and that engaging in dialogic feedback may better enable participant engagement in the self-regulation cycle of reflection, planning and intra-task monitoring.

10.7 PERSONAL AND PROFESSIONAL REFLECTION

This thesis is the summation of four years of work and this research degree has been the largest and most challenging academic activity that I have ever undertaken. At times I have felt complete cognitive overload, with my intellectual ability stretched to its limit by the complex critique and original thought that has been required. This is tempered by a real sense of privilege at the opportunity to contribute something original and potentially useful to medical education.

This research project has presented me with a challenge unlike any other I have experienced before and is the best example I now have of delayed gratification. In clinical life, particularly in surgery, our workload is split into well-defined segments. These may be based on time periods or relate to set volumes of work but satisfaction is felt on a regular basis, when each and any of these tasks are completed. The combination of size and the complexity and of a research degree, with each part related and reliant on the others, does not lend itself to well-defined incremental gains with palpable reward. The hard-yards of thinking and rethinking, writing and rewriting can only be motivated by a belief that the hard work would be worth it in the end. Even reflecting on that now it’s not quite possible to feel that gratification yet; it is difficult to imagine a time after this thesis! The independent nature of the work is not like clinical life either; sometimes I have felt a bit lonely but never alone. During periods of particular inertia, I have woken up early or found it difficult to sleep because of anxious thoughts. I don’t think there has ever been a time that I genuinely thought I wouldn't finish this degree but definitely times when I didn't understand where the hours it required were going to come from.

This degree has been a cause for personal change and, I hope, improved awareness if not insight. I have had to deal with sometimes feeling overwhelmed, not knowing the best way forward and having to take action anyway. I have had to learn to manage the internal tension caused by having no choice but to invest in time thinking when the urge is to produce something tangible in order to sense progress. At times I have had to be selfish with my time; missing out on spending time with my partner and family and justifying this with having no choice, it being for the greater good or for the sake of future happiness. The truth is that ultimately, I chose to do this degree, alongside a busy clinical job, for my own gain and, therefore, I have to take responsibility for the added stress that it has placed on my relationships. I think of all the effects associated with undertaking this degree, this is the only one that still makes me uncomfortable.

Whilst the discomfort described above might be summarised as the costs of deciding to do a research degree, I am under no doubt as to the benefits to me as a clinician, maturing researcher, educator and learner. Spending some time part-immersed in the academic world of medical education has exposed me to people and perspectives very different from my clinical colleagues. It has broadened my horizons and made me consider my work differently, approaching clinical problems and colleagues with less tunnel-vision. It has made me appreciate the value of an academic perspective and how important it is to think and listen when trying to innovate and improve. It has made me more aware of the power but also the limitations of a clinician's drive to make decisions and implement rather than muse.

I still find it difficult to consider myself as a researcher; I am clear than it's not a role that I am naturally good at. I think these feelings of inadequacy have sometimes been a barrier to engaging with my supervisors. However, when I consider the immaturity of my skills and knowledge at the beginning of this degree and what they are now, I know progress has been made. I am grateful to my supervisors for leading me through this process and I learnt a lot about academia and supervision from reflecting on how they have managed and worked with me. There are many specific research skills that I have gained from this degree – structured literature review, quantitative study design, managing a large-scale experimental project – but the most valuable change for me has been the shift in perspective from a quantitative surgeon who is concerned only with the ‘what’, to a qualitatively literate researcher who understands the importance of the ‘why’.

Interestingly for me, this research degree has had an unsettling effect on how I consider myself as a teacher. Previously, I would have considered myself a good teacher and I would take satisfaction in my ability to convey information to undergraduate students and junior colleagues. However, reviewing the literature theorising feedback, my experience of undertaking this project and analysing the results it has produced has convinced me that an effective educator is a facilitator not a director; a king-maker but not a king. It has impressed upon me the limited value of adopting the ego-boosting role of the learned tutor who has all of the answers and has made me much more supportive of promoting an active learner role and the consequential learner realisation of self-efficacy.

Finally, and potentially most importantly, I think this project has helped me as a learner. In relation to the fine grain, it has given me a fresh perspective on how I learn and how I can use feedback. This is helpful and relevant to every day of my clinical work. With respect to the large grain, this thesis has taught me the value of perseverance. It is the hardest thing I have ever done and the reward is a greater sense of competence and self-efficacy. I hope it is a stepping stone for future learning and self-improvement.

GLOSSARY

Ceiling effect	A ceiling effect occurs when a high proportion of subjects in a study have maximum scores on the observed variable. This makes discrimination among subjects among the top end of the scale impossible.
Cognitive load	The total amount of mental activity imposed on working memory in any one instant.
Confirmability	The degree to which study findings which are based on the viewpoints of participant.
Credibility	Confidence in the truth of the data and interpretations of them.
Dependability	The stability of data over time and over conditions; an evaluation of the quality of the integrated processes of data collection, data analysis, and theory generation.
Dialogic feedback	A feedback model based upon co-constucted discussion between tutor and learner.
Floor effect	A floor effect occurs when a measure possesses a distinct lower limit for potential responses and a large concentration of participants score at or near this limit (the opposite of a ceiling effect). This makes discrimination among subjects among the bottom end of the scale impossible.
Generalisability	The extent to which research findings can be applied to settings other than that in which they were originally tested.
Information transfer feedback	A linear model of feedback, in which information is conveyed from tutor to learner.
Insider researcher	A researcher plays an important role in the topic being researched
Measure of performance	A quantification of performance.
Psychomotor task	A physical skill or action reliant upon cognition and understanding for completion.
Reflexivity	An attitude of attending systematically to the context of knowledge construction, especially to the effect of the researcher, at every step of the research process.
Reliability	The ability of a test to measure consistently.
Saturation (of data)	A phenomenon describing the point at which no new data is contributed to the main category during content analysis during qualitative data analysis.
Self-efficacy	One's belief in one's ability to succeed in specific situations or accomplish a task.

Self-directed learning	A process through which learners identify missing knowledge and engage with learning materials in a self-determined way to fill them.
Self-regulated learning	A process of taking control of and evaluating one's own learning and behaviour, characterised by metacognition, strategic action, and motivation to learn.
Self-regulation of task	Self-generated thoughts, feelings and actions that are planned and cyclically adapted to the attainment of personal goals, characterised by three phases: forethought, performance and self-reflection.
Transferability	How well study findings can be transferred to other settings, contexts or groups.
Triangulation	The use of different methods in order to check the validity of the study findings and to minimise source of errors.
Validity	The extent to which an instrument measures what it is intended to measure.

BIBLIOGRAPHY

Ajjawi, R. (2012) Going beyond 'received and understood' as a way of conceptualising feedback. Med Educ, 46(10), 1018-9.

Archer, J. (2013) Feedback: it's all in the CHAT. Med Educ, 47(11), 1059-61.

Archer, J. C. (2010) State of the science in health professional education: effective feedback. Med Educ, 44(1), 101-8.

Aronson, L., Niehaus, B., Hill-Sakurai, L., Lai, C. & O'Sullivan, P. S. (2012) A comparison of two methods of teaching reflective ability in Year 3 medical students. Med Educ, 46(8), 807-14.

Artino, A. R., Jr., Cleary, T. J., Dong, T., Hemmer, P. A. & Durning, S. J. (2014) Exploring clinical reasoning in novices: a self-regulated learning microanalytic assessment approach. Med Educ, 48(3), 280-91.

Attia, M. & Edge, J. (2017) Be(com)ing a reflexive researcher: a developmental approach to research methodology, Open Review of Educational Research, 4(1), 33-45.

Backstein, D., Agnidis, Z., Regehr, G. & Reznick, R. (2004) The effectiveness of video feedback in the acquisition of orthopedic technical skills. Am J Surg, 187(3), 427-32.

Bing-You, R. G. & Trowbridge, R. L. (2009) Why medical educators may be failing at feedback. Jama, 302(12), 1330-1.

Boehler, M. L., Rogers, D. A., Schwind, C. J., Mayforth, R., Quin, J., Williams, R. G. & Dunnington, G. (2006) An investigation of medical student reactions to feedback: a randomised controlled trial. Med Educ, 40(8), 746-9.

Boud, D. & Molloy, E. (2013) Rethinking models of feedback for learning: the challenge of design. Assessment & Evaluation in Higher Education, 38(6), 698-712.

Bounds, R., Bush, C., Aghera, A., Rodriguez, N., Stansfield, R. B. & Santen, S. A. (2013) Emergency medicine residents' self-assessments play a critical role when receiving feedback. Acad Emerg Med, 20(10), 1055-61.

Boyle, E., Al-Akash, M., Gallagher, A. G., Traynor, O., Hill, A. D. & Neary, P. C. (2011a) Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure. Postgrad Med J, 87(1030), 524-8.

Boyle, E., O'Keeffe, D. A., Naughton, P. A., Hill, A. D., McDonnell, C. O. & Moneley, D. (2011b) The importance of expert feedback during endovascular simulator training. J Vasc Surg, 54(1), 240-248 e1.

Branch, W. T., Jr. & Paranjape, A. (2002) Feedback and reflection: teaching methods for clinical settings. Acad Med, 77(12 Pt 1), 1185-8.

Brydges, R. & Butler, D. (2012) A reflective analysis of medical education research on self-regulation in learning and practice. Med Educ, 46(1), 71-9.

Brydges, R., Nair, P., Ma, I., Shanks, D. & Hatala, R. (2012) Directed self-regulated learning versus instructor-regulated learning in simulation training. Med Educ, 46(7), 648-56.

Burr, S. A., Brodier, E. & Wilkinson, S. (2013) Delivery and use of individualised feedback in large class medical teaching. BMC Med Educ, 13, 63.

Butler, D. L. & Brydges, R. (2013) Learning in the health professions: what does self-regulation have to do with it? Med Educ, 47(11), 1057-9.

Carless, D. (2006) Differing perceptions in the feedback process. Studies in higher education, 31(2), 219-233.

Carless, D., Salter, D., Yang, M. & Lam, J. (2011) Developing sustainable feedback practices. Studies in Higher Education, 36(4), 395-407.

Carrier, A. & Morin, C. (2014) Enabling students' self-regulation and teachers' feedback in concept mapping. Med Educ, 48(5), 523-4.

Chang, A., Chou, C. L., Teherani, A. & Hauer, K. E. (2011) Clinical skills-related learning goals of senior medical students after performance feedback. Med Educ, 45(9), 878-85.

Clay, A. S., Que, L., Petrusa, E. R., Sebastian, M. & Govert, J. (2007) Debriefing in the intensive care unit: a feedback tool to facilitate bedside teaching. Crit Care Med, 35(3), 738-54.

Cleary, T. J. & Sandars, J. (2011) Assessing self-regulatory processes during clinical skill performance: a pilot study. Med Teach, 33(7), e368-74.

Cook, M. R., Watters, J. M., Barton, J. S., Kamin, C., Brown, S. N., Deveney, K. E. & Kiraly, L. N. (2015) A flexible postoperative debriefing process can effectively provide formative resident feedback. J Am Coll Surg, 220(5), 959-67.

Corrigan, G. (2012) Self-regulated learning in medical education: the next steps. Med Educ, 46(9), 920.

Devi, V., Mandal, T., Kodidela, S. & Pallath, V. (2012) Integrating students' reflection-in-learning and examination performance as a method for providing educational feedback. J Postgrad Med, 58(4), 270-4.

El Saadawi, G. M., Azevedo, R., Castine, M., Payne, V., Medvedeva, O., Tseytlin, E., Legowski, E., Jukic, D. & Crowley, R. S. (2010) Factors affecting feeling-of-knowing in a medical intelligent tutoring system: the role of immediate feedback as a metacognitive scaffold. Adv Health Sci Educ Theory Pract, 15(1), 9-30.

Elnicki, D. M. & Zalenski, D. (2013) Integrating medical students' goals, self-assessment and preceptor feedback in an ambulatory clerkship. Teach Learn Med, 25(4), 285-91.

Ende, J. (1983) Feedback in clinical medical education. Jama, 250(6), 777-81.

Eva, K. W., Munoz, J., Hanson, M. D., Walsh, A. & Wakefield, J. (2010) Which factors, personal or external, most influence students' generation of learning goals? Acad Med, 85(10 Suppl), S102-5.

Farjad Sultan, S., Iohom, G. & Shorten, G. (2013) Effect of feedback content on novices' learning ultrasound guided interventional procedures. Minerva Anestesiol, 79(11), 1269-80.

Farquharson, A. L., Cresswell, A. C., Beard, J. D. & Chan, P. (2013) Randomized trial of the effect of video feedback on the acquisition of surgical skills. Br J Surg, 100(11), 1448-53.

Finn, K. M., Roy, C. L. & Katz, J. T. (2007) Improving the quality of intern documentation through structured feedback. Med Educ, 41(11), 1101-2.

Ghaderi, I., Auvergne, L., Park, Y. S. & Farrell, T. M. (2015) Quantitative and qualitative analysis of performance during advanced laparoscopic fellowship: a curriculum based on structured assessment and feedback. Am J Surg, 209(1), 71-8.

Gonzalez, R., Bowers, S. P., Smith, C. D. & Ramshaw, B. J. (2004) Does setting specific goals and providing feedback during training result in better acquisition of laparoscopic skills? Am Surg, 70(1), 35-9.

Grantcharov, T. P., Schulze, S. & Kristiansen, V. B. (2007) The impact of objective assessment and constructive feedback on improvement of laparoscopic performance in the operating room. Surg Endosc, 21(12), 2240-3.

Hattie, J. & Timperley, H. (2007) The power of feedback. Review of educational research, 77(1), 81-112.

Hawkins, S. C., Osborne, A., Schofield, S. J., Pournaras, D. J. & Chester, J. F. (2012) Improving the accuracy of self-assessment of practical clinical skills using video feedback--the importance of including benchmarks. Med Teach, 34(4), 279-84.

Hettema, J. E., Ratanawongsa, N., Manuel, J. K., Ciccarone, D., Coffa, D., Jain, S. & Lum, P. J. (2012) A SBIRT curriculum for medical residents: development of a performance feedback tool to build learner confidence. Subst Abus, 33(3), 241-50.

Hoffman, B. D. (2015) Using self-determination theory to improve residency training: learning to make omelets without breaking eggs. Acad Med, 90(4), 408-10.

Hoogenes, J., Mironova, P., Safir, O., McQueen, S. A., Abdelbary, H., Drexler, M., Nousiainen, M., Ferguson, P., Kraemer, W., Alman, B., Reznick, R. K. & Sonnadara, R. R. (2015) Student-led learning: a new teaching paradigm for surgical skills. Am J Surg, 209(1), 107-14.

Judkins, T. N., Oleynikov, D. & Stergiou, N. (2006) Real-time augmented feedback benefits robotic laparoscopic training. Stud Health Technol Inform, 119, 243-8.

Kannappan, A., Yip, D. T., Lodhia, N. A., Morton, J. & Lau, J. N. (2012) The effect of positive and negative verbal feedback on surgical skills performance and motivation. J Surg Educ, 69(6), 798-801.

Kaul, P., Gong, J. & Guiton, G. (2014) Effective feedback strategies for teaching in pediatric and adolescent gynecology. J Pediatr Adolesc Gynecol, 27(4), 188-93.

Kluger, A. N. & Van Dijk, D. (2010) Feedback, the various tasks of the doctor, and the feedforward alternative. Med Educ, 44(12), 1166-74.

Kraut, A., Yarris, L. M. & Sargeant, J. (2015) Feedback: Cultivating a Positive Culture. J Grad Med Educ, 7(2), 262-4.

Kruglikova, I., Grantcharov, T. P., Drewes, A. M. & Funch-Jensen, P. (2010) The impact of constructive feedback on training in gastrointestinal endoscopy using high-fidelity Virtual-Reality simulation: a randomised controlled trial. Gut, 59(2), 181-5.

Li, Q., Ma, E. L., Liu, J., Fang, L. Q. & Xia, T. (2011) Pre-training evaluation and feedback improve medical students' skills in basic life support. Med Teach, 33(10), e549-55.

Lipsett, P. A., Harris, I. & Downing, S. (2011) Resident self-other assessor agreement: influence of assessor, competency, and performance level. Archives of Surgery, 146(8), 901-906.

Malterud, K. (2001). Qualitative research: Standards, challenges and guidelines. Lancet, 358, 483-488.

Maxwell, J. (2010) Using Numbers in Qualitative Research. Qual Inq, 16(6), 475–482.

Milan, F. B., Parish, S. J. & Reichgott, M. J. (2006) A model for educational feedback based on clinical communication skills strategies: beyond the "feedback sandwich". Teach Learn Med, 18(1), 42-7.

Molloy, E. K. (2010) The feedforward mechanism: a way forward in clinical learning? Med Educ, 44(12), 1157-8.

Murad, M. H., Chu, H., Lin, L., Wang, Z. (2018) BMJ Evidence-Based Medicine, 23, 84–86.

Murdoch-Eaton, D. (2012) Feedback: the complexity of self-perception and the transition from 'transmit' to 'received and understood'. Med Educ, 46(6), 538-40.

Nicol, D. Principles of good assessment and feedback: Theory and practice 2007.

Norcini, J. (2010) The power of feedback. Med Educ, 44(1), 16-7.

Nothnagle, M., Reis, S., Goldman, R. & Diemers, A. (2010) Development of the GPSE: a tool to improve feedback on procedural skills in residency. Fam Med, 42(7), 507-13.

O'Connor, A., Schwaitzberg, S. D. & Cao, C. G. (2008) How much feedback is necessary for learning to suture? Surg Endosc, 22(7), 1614-9.

Oestergaard, J., Bjerrum, F., Maagaard, M., Winkel, P., Larsen, C. R., Ringsted, C., Gluud, C., Grantcharov, T., Ottesen, B. & Soerensen, J. L. (2012) Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized educational trial. BMC Med Educ, 12, 7.

Overeem, K. (2010) 'Paying it forward': performance improvement through feedforward interviews. Med Educ, 44(12), 1159-61.

Paschold, M., Huber, T., Zeissig, S. R., Lang, H. & Kneist, W. (2014) Tailored instructor feedback leads to more effective virtual-reality laparoscopic training. Surg Endosc, 28(3), 967-73.

Pope, C., Zeibland, S., & Mays, N. (2000) Analysing qualitative data. BMJ, 320, 114–6.

Porte, M. C., Xeroulis, G., Reznick, R. K. & Dubrowski, A. (2007) Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. Am J Surg, 193(1), 105-10.

Pusic, M. V., Brydges, R., Kessler, D., Szyld, D., Nachbar, M. & Kalet, A. (2014) What's your best time? Chronometry in the learning of medical procedures. Med Educ, 48(5), 479-88.

Rafiq, A., Tamariz, F., Boanca, C., Lavrentyev, V. & Merrell, R. C. (2008) Objective assessment of training surgical skills using simulated tissue interface with real-time feedback. J Surg Educ, 65(4), 270-4.

Riordan, T. & Loacker, G. (2009) Collaborative and Systemic Assessment of Student Learning: From Principles to Practice, in Joughin, G. (Ed), Assessment, Learning and Judgement in Higher Education. Dordrecht: Springer Netherlands, 1-18.

Ritchie, J. & Spencer, L. (1994) Qualitative data analysis for applied policy research, in Bryman, A. a. B., RG (ed), Analyzing qualitative data. London: Francis & Taylor Group, 173-194.

Rizan, C., Elsey, C., Lemon, T., Grant, A. & Monrouxe, L. V. (2014) Feedback in action within bedside teaching encounters: a video ethnographic study. Med Educ, 48(9), 902-20.

Rogers, D. A., Regehr, G., Howdieshell, T. R., Yeh, K. A. & Palm, E. (2000) The impact of external feedback on computer-assisted learning for surgical technical skill training. Am J Surg, 179(4), 341-3.

Rolfe, I. & McPherson, J. (1995) Formative assessment: how am I doing? Lancet, 345(8953), 837-9.

Rudland, J., Wilkinson, T., Wearn, A., Nicol, P., Tunny, T., Owen, C. & O'Keefe, M. (2013) A student-centred feedback model for educators. Clin Teach, 10(2), 99-102.

Sadideen, H. & Kneebone, R. (2012) Practical skills teaching in contemporary surgical education: how can educational theory be applied to promote effective learning? Am J Surg, 204(3), 396-401.

Sagasser, M. H., Kramer, A. W. & van der Vleuten, C. P. (2012) How do postgraduate GP trainees regulate their learning and what helps and hinders them? A qualitative study. BMC Med Educ, 12, 67.

Sandars, J. (2012) Future direction for research in self-regulated learning in medical education. Med Educ, 46(6), 626.

Sandars, J. (2015) The challenge of feedback-insights from non-medical educational research. Int J Med Educ, 6, 1-3.

Sandars, J. & Cleary, T. J. (2011) Self-regulation theory: applications to medical education: AMEE Guide No. 58. Med Teach, 33(11), 875-86.

Sandelowski, M. (2001) Real qualitative researchers do not count: the use of numbers in qualitative research. Res Nurs Health, 24(3), 230-40.

Sargeant, J., Mann, K., van der Vleuten, C. & Metsemakers, J. (2008) "Directed" self-assessment: practice and feedback within a social context. J Contin Educ Health Prof, 28(1), 47-54.

Schartel, S. A. (2012) Giving feedback - an integral part of education. Best Pract Res Clin Anaesthesiol, 26(1), 77-87.

Schum, T. R., Krippendorf, R. L. & Biernat, K. A. (2003) Simple feedback notes enhance specificity of feedback to learners. Ambul Pediatr, 3(1), 9-11.

Shanks, D., Brydges, R., den Brok, W., Nair, P. & Hatala, R. (2013) Are two heads better than one? Comparing dyad and self-regulated learning in simulation training. Med Educ, 47(12), 1215-22.

Sobral, D. T. (2005) Medical students' mindset for reflective learning: a revalidation study of the reflection-in-learning scale. Adv Health Sci Educ Theory Pract, 10(4), 303-14.

Spanager, L., Dieckmann, P., Beier-Holgersen, R., Rosenberg, J. & Oestergaard, D. (2015) Comprehensive feedback on trainee surgeons' non-technical skills. Int J Med Educ, 6, 4-11.

Strandbygaard, J., Bjerrum, F., Maagaard, M., Winkel, P., Larsen, C. R., Ringsted, C., Gluud, C., Grantcharov, T., Ottesen, B. & Sorensen, J. L. (2013) Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial. Ann Surg, 257(5), 839-44.

Tavakol, M. & Sandars, J. (2014) Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part II. Med Teach, 36: 746–756

Tavakol, M. & Sandars, J. (2014) Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part II. Med Teach, 36:10, 838-848.Telio, S., Ajjawi, R. & Regehr, G. (2015) The "educational alliance" as a framework for reconceptualizing feedback in medical education. Acad Med, 90(5), 609-14.

Turan, S. & Konan, A. (2012) Self-regulated learning strategies used in surgical clerkship and the relationship with clinical achievement. J Surg Educ, 69(2), 218-25.

van de Ridder, J. M., Stokking, K. M., McGaghie, W. C. & ten Cate, O. T. (2008) What is feedback in clinical education? Med Educ, 42(2), 189-97.

Van Sickle, K. R., Gallagher, A. G. & Smith, C. D. (2007) The effect of escalating feedback on the acquisition of psychomotor skills for laparoscopy. Surg Endosc, 21(2), 220-4.

Veloski, J., Boex, J. R., Grasberger, M. J., Evans, A. & Wolfson, D. B. (2006) Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. Medical teacher, 28(2), 117-128.

Watling, C., Driessen, E., van der Vleuten, C. P., Vanstone, M. & Lingard, L. (2012) Understanding responses to feedback: the potential and limitations of regulatory focus theory. Med Educ, 46(6), 593-603.

Winne, P. H. & Perry, N. E. (2000) Chapter 16 - Measuring Self-Regulated Learning A2 - Boekaerts, Monique, in Pintrich, P. R. & Zeidner, M. (Ed), Handbook of Self-Regulation. San Diego: Academic Press, 531-566.

Wojcikowski, K. & Kirk, L. (2013) Immediate detailed feedback to test-enhanced learning: an effective online educational tool. Med Teach, 35(11), 915-9.

Wulf, G., Shea, C. & Lewthwaite, R. (2010) Motor skill learning and performance: a review of influential factors. Med Educ, 44(1), 75-84.

Xeroulis, G. J., Park, J., Moulton, C. A., Reznick, R. K., Leblanc, V. & Dubrowski, A. (2007) Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. Surgery, 141(4), 442-9.

Zimmerman, B. (1996) Enhancing student academic and health functioning: A self-regulatory perspective. School Psychology Quarterly, 11(1), 47-66.

Zimmerman, B. (2000) Attaining Self-Regulation: A Social Cognitive Perspective, in Monique Boekaerts, P. R. P. a. M. Z. (Ed), Handbook of Self-Regulation. New York: Academic Press, 13-39.

Zimmerman, B. & Moylan, A. (2009) Self-Regulation: Where Metacognition and Motivation Intersect. Dalam Hacker, D.J. (Ed), Handbook of Metacognition in Education. New York: Routledge, 299-316.

Zuberi, R. W. (2012) Layers within layers ... self-regulation in a complex learning environment. Med Educ, 46(1), 7-8.

APPENDICES

Appendix A: Ethical approval UREC 14134

UNIVERSITY OF DUNDEE

School of Psychology

University of Dundee Research Ethics Committee

University of Dundee
Dundee
DD1 4HN

29th October 2014

Dear Dr Gill,

Application Number: UREC 14134

Title: Feedback and learning in surgical education

I am writing to you to advise you that your ethics application has been reviewed and approved by the University of Dundee Research Ethics Committee.

Approval is valid for three years from the date of this letter. Should your study continue beyond this point, please request a renewal of the approval.

Any changes to the approved documentation (e.g., study protocol, information sheet, consent form), must be approved by UREC.

Yours sincerely,

Dr Astrid Schloerscheidt
Chair, University of Dundee Research Ethics Committee

UNIVERSITY OF DUNDEE Dundee DD1 4HN Scotland UK *t* +44(0)1382 229993
e psych@dundee.ac.uk www.dundee.ac.uk/psychology

Appendix B: Pilot one participant information sheet

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

PILOT STUDY PARTICIPANT INFORMATION SHEET

INVITATION TO TAKE PART IN A RESEARCH STUDY
You are being asked to take part in a pilot for a research study, which aims to investigate the possible relationship between feedback and surgical skill development. The lead investigator is Miss Sarah Gill, a surgical trainee in the East of Scotland Deanery, who is undertaking this study as part of an MD in Medical Education. The study is being supervised by Dr Rola Ajjawi, a Senior Lecturer in Medical Education, interested in workplace learning and feedback.

Please let me stress that your participation is completely voluntary. Your performance is confidential and it will in no way influence your grading at Dundee Medical School.

WHAT IT INVOLVES & THE PURPOSE OF THE RESEARCH STUDY
Participation in this study will involve performing a surgical task on a laparoscopic simulator and, in some cases, participation in a feedback session in relation to performance. This process will be repeated over several times during the session. The task will be explained prior to you undertaking it and your performance will be measured but this will remain confidential between only you and the investigator. All data will be stored anonymously. The level at which you are able to repeat the task is not, in isolation, what is being studied. The relationship between feedback and the trend of performance is what is of interest. It is not possible to "fail" in the study as it is not a test.

It is proposed that investigation into this field and your participation in this research would benefit future surgical trainees. We hope that the participants in this study will find the experience interesting and will benefit from the surgical skills acquired as part of the study.

TIME COMMITMENT
The session will last approximately 30 minutes.

COST, REIMBURSEMENT AND COMPENSATION
Your participation in this study is voluntary and of course participating in the study does not cost anything. You will be awarded a certificate of participation for your portfolio as thanks for taking part in the study.

RISKS AND TERMINATION OF PARTICIPATION
There are no known risks for you in this study. You may decide to stop being a part of the research study at any time without explanation and without penalty.

CONFIDENTIALITY/ANONYMITY
The data collected will include demographics, such as age and gender, and experimental results but no one will be able to link this data to your identity and name. The data will be seen only by the researchers and will not be made available to anyone else. The results will be published as part of an MD thesis and research papers.

FOR FURTHER INFORMATION ABOUT THIS RESEARCH STUDY
Sarah Gill will be glad to answer your questions about this study at any time and can be contacted using the details below. Should you wish to find out about the final results of this study, she will be happy to take your contact details and send you a copy of the results.

Address:	Medical Education Directorate, Level 8, Dundee Medical School, Ninewells Hospital, Dundee, DD1 9SY
Email:	s.v.gill@dundee.ac.uk
Telephone:	01382 383528

Appendix C: Pilot one participant consent form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

PILOT STUDY PARTICIPANT CONSENT FORM

This research study aims to investigate the possible relationship between feedback and surgical skill development. Participation in this study will involve performing a surgical task on a laparoscopic simulator and, in some cases, participation in a feedback session in relation to performance.

By signing below you are indicating that you have read and understood the Participant Information Sheet and that you agree to take part in this research study.

Signed (participant): …………………………………………………………… **Date:** …………………………………………

Name (participant): ……………………………………………………………

Signed investigator): …………………………………………………………… **Date:** …………………………………………

Name (investigator): Miss Sarah Gill

Appendix D: Pilot one pre-study data form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study pre-task data collection

Participant number:	
Sex:	
Age:	
Hand dominance:	

Pre-study **confidence** level:

- Please rate below your confidence with laparoscopic tasks on the scale 1 – 5
- A score of 1 reflects the least confidence
- A score of 5 reflects the most confident

1 2 3 4 5

Pre-study **experience** level:

- Please rate below your experience with laparoscopic work/practice on the scale 1 – 5
- A score of 1 reflects the little or no experience
- A score of 5 reflects a high level of experience

1 2 3 4 5

Please describe how much you practice you have had in using any laparoscopic training instruments:

Eg. 30 minutes using laparoscopic training machine in Cushieri skills centre

Pre-study **trainer confidence** level:

- Please rate below your confidence in the trainer in their ability to provide feedback using the scale 1 – 5
- A score of 1 reflects the little or no confidence
- A score of 5 reflects a high level of confidence

1 2 3 4 5

Appendix E: Pilot one task instructions sheet

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study: Task instruction sheet

Aim of the task:

This study features a simulated laparoscopic task. The aim of the task on the *LapSim* is to correctly apply two clips across the vessel lying in the abdominal cavity, before then dividing the vessel. Any dropped clips plus any large volumes of blood must be removed before the task is deemed completed.

Optimal task completion: Upon initially grasping the vessel, two green 'clip target areas' will appear. Ideally, a clip should be placed in each of these two zones. Once the clips have been applied in the target areas, a blue 'cutting target area' will appear; the vessel should be divided in this area.

Adequate task completion: The clips can be applied at any two points of the vessel and, as long as they're adequately applied, they will be functional and prevent bleeding of the divided vessel. Equally, the vessel can be divided anywhere between the two functioning clips.

Instrumentation:

You have a laparoscopic tool for each hand, left and right. Four types of instruments are available for you to use:

- Grasper
- Clip applier
- Curved scissors
- Suction

The foot pedal must be pressed to activate the suction device.
You can change the instrument used in either hand at any time: Pull the instrument back as far as you can (as if removing it from the abdomen); a menu of the four instruments will appear on the screen with the selected instrument highlighted. Click together the handles of the instrument on that side to cycle between the instruments and stop when the one you want to use is highlighted, then re-insert the instrument.
You can change the angle of the end of the instrument by using cogwheel by your fingertips.

Pitfalls and complications:

<u>*Stretch damage and bleeding:*</u> Blood vessels are susceptible to stretch damage. If the blood vessel is stretched, it will start to turn red and it can bleed, requiring a clip to be placed below the bleeding point to control it. If the blood vessel is stretched beyond its limit, it will break. The task can be completed by placing a clip on each end of the vessel to stop the bleeding. Minimise pressure on the abdominal wall from the instrument tips as this can cause damage. If the tips of the instrument are pressed into the abdominal wall, the screen will go red. This disappears once the pressure is relieved.

<u>*Dropped clips:*</u> If the jaws of the clip applier are partially closed prior to being applied to the vessel, the clip will not spring open again and it needs to be discarded by dropping it. Fully close the jaws of the clip applier and the clip will drop to the abdominal floor. As previously stated, all dropped clips must be recovered with the graspers.

Clip placement: For the clips to be effective, it needs to be placed across the vessel with equal length of clip showing on either side. If the clip isn't placed correctly, it will not occlude the vessel properly and the clip won't function, meaning the vessel will bleed if divided. If this occurs, simply ignore this clip and place another. These clips do not have to be removed before the end of the procedure.

At the end of the task:

Once the task has been completed, the screen will simply fade out.

Appendix F: Pilot one information transfer feedback form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study: Information transfer group
Feedback checklist

Participant number: Task number: Time taken for FB:

	Component	**Rating**			**Comments**
G1	Follows and agreed, logical sequence or protocol for the procedure	N	D	S	
G2	Consistently handles tissue well with minimal damage	N	D	S	
G3	Controls bleeding promptly	N	D	S	
G4	Uses instruments appropriately and safely	N	D	S	
G5	Proceeds at an appropriate pace with economy of movement	N	D	S	
G6	Deals calmly and effectively with unexpected events and complications	N	D	S	
T1	Grasps vessel carefully with grasper to allow clip application	N	D	S	
T2	Applies clips correctly to either side of vessel	N	D	S	
T3	Cuts vessel safely with scissors	N	D	S	
T4	Retrieves dropped clips	N	D	S	

Level of performed task		
0	Unable to complete task	
1	Able to complete task with significant difficulty in several components	
2	Able to complete task with mild-moderate difficulty in several components	
3	Able to perform all elements of task with mild difficulty in only some components	
4	Able to perform all elements of task, exhibiting high skill level	

Appendix G: Laparoscopic procedure based assessment

ISCP INTERCOLLEGIATE SURGICAL CURRICULUM PROGRAMME

Procedure Based Assessment (PBA)

Procedure name ELECTIVE - Generic Laparoscopic Hernia Repair [General Surgery]

Ratings

Your ratings should be judged against the standard for the Certification. Assessors are normally consultants (senior trainees may be assessors depending upon their training level and the complexity of the procedure) IMPORTANT: The trainee should explain what he/she intends to do throughout the procedure. The Assessor should provide verbal prompts if required, and intervene if patient safety is at risk. A rating of Satisfactory can only be given if no prompting or intervention was required

Intra-operative technique: global (G) and task-specific items (T)

IT1(G)	Follows an agreed, logical sequence or protocol for the procedure	N	D	S
IT2(G)	Consistently handles tissue well with minimal damage	N	D	S
IT3(G)	Controls bleeding promptly by an appropriate method	N	D	S
IT4(G)	Demonstrates a sound technique of knots and sutures/staples	N	D	S
IT5(G)	Uses instruments appropriately and safely	N	D	S
IT6(G)	Proceeds at appropriate pace with economy of movement	N	D	S
IT7(G)	Anticipates and responds appropriately to variation e.g. anatomy	N	D	S
IT8(G)	Deals calmly and effectively with unexpected events/complications	N	D	S
IT9(G)	Uses assistant(s) to the best advantage at all times	N	D	S
IT10(G)	Communicates clearly and consistently with the scrub team	N	D	S
IT11(G)	Communicates clearly and consistently with the anaesthetist	N	D	S
IT12(T)	Establishes safe and appropriate laparoscopic access – camera port, insufflation and ergonomic instrument port placement	N	D	S
IT13(T)	Identifies and prepares hernial defect by division of adhesions and preparation of site for mesh placement	N	D	S
IT14(T)	Fixes mesh in place using an appropriate technique	N	D	S

Appendix H: Dialogic feedback form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study: Dialogic group
Feedback model

Participant number: Task number: Time for FB:

	Component	**Systematic / tactical awareness**	**Technical skills**
G1	Logical sequencing		
G2	Tissue handling		
G3	Use of instruments		
G4	Pace with economy of movement		
G5	Deals calmly and effectively with unexpected events and complications		
G6	Evidence of self-regulation		
T1	Use of grasper to allow clip application		
T2	Applies clips correctly to either side of vessel		
T3	Control of bleeding		
T4	Cuts vessel safely with scissors		
T5	Minimises dropped clips / method retrieval		

Appendix I: Pilot two participant information sheet

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

PILOT STUDY PARTICIPANT INFORMATION SHEET

INVITATION TO TAKE PART IN A RESEARCH STUDY
You are being asked to take part in a pilot for a research study, which aims to investigate the possible relationship between feedback and surgical skill development. The lead investigator is Miss Sarah Gill, a surgical trainee in the East of Scotland Deanery, who is undertaking this study as part of an MD in Medical Education. The study is being supervised by Dr Rola Ajjawi, a Senior Lecturer in Medical Education, interested in workplace learning and feedback.

Please let me stress that your participation is completely voluntary. Your performance is confidential and it will in no way influence your grading at Dundee Medical School.

WHAT IT INVOLVES & THE PURPOSE OF THE RESEARCH STUDY
Participation in this study will involve performing surgical tasks on a laparoscopic simulator and participation in a feedback session in relation to performance. This process will be repeated over several times during the sessions. The task will be explained prior to you undertaking it and your performance will be measured but this will remain confidential between only you and the investigator. All data will be stored anonymously. The level at which you are able to repeat the task is not, in isolation, what is being studied. The relationship between feedback and the trend of performance is what is of interest. It is not possible to "fail" in the study as it is not a test.

It is proposed that investigation into this field and your participation in this research would benefit future surgical trainees. We hope that the participants in this study will find the experience interesting and will benefit from the surgical skills acquired as part of the study.

TIME COMMITMENT
The first session will last approximately 45 minutes and you are asked to attend a follow up session in 1 weeks' time.

COST, REIMBURSEMENT AND COMPENSATION
Your participation in this study is voluntary and of course participating in the study does not cost anything. You will be awarded a certificate of participation for your portfolio as thanks for taking part in the study.

RISKS AND TERMINATION OF PARTICIPATION
There are no known risks for you in this study. You may decide to stop being a part of the research study at any time without explanation and without penalty.

CONFIDENTIALITY/ANONYMITY
The data collected will include demographics, such as age and gender, and experimental results but no one will be able to link this data to your identity and name. The data will be seen only by the researchers and will not be made available to anyone else. The results will be published as part of an MD thesis and research papers.

FOR FURTHER INFORMATION ABOUT THIS RESEARCH STUDY
Sarah Gill will be glad to answer your questions about this study at any time and can be contacted using the details below. Should you wish to find out about the final results of this study, she will be happy to take your contact details and send you a copy of the results.

Address: Medical Education Directorate, Level 8, Dundee Medical School, Ninewells Hospital, Dundee, DD1 9SY
Email: s.v.gill@dundee.ac.uk
Telephone: 01382 383528

Appendix J: Pilot two participant consent form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

PILOT STUDY PARTICIPANT CONSENT FORM

This research study aims to investigate the possible relationship between feedback and surgical skill development. Participation in this study will involve performing a surgical task on a laparoscopic simulator and participation in a feedback session in relation to performance.

By signing below you are indicating that you have read and understood the Participant Information Sheet and that you agree to take part in this research study.

Signed (participant): ………………………………………………………… **Date:** ………………………………………

Name (participant): …………………………………………………………

Signed investigator): ………………………………………………………… **Date:** ………………………………………

Name (investigator): Miss Sarah Gill

Appendix K: Pilot two pre-study data form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study pre-task data collection

Participant number:	
Sex:	
Age:	
Hand dominance:	

Pre-study **confidence** level:

- Please rate below your confidence with laparoscopic tasks on the scale 1 – 5
- A score of 1 reflects the least confidence
- A score of 5 reflects the most confident

1 2 3 4 5

Pre-study **experience** level:

- Please rate below your experience with laparoscopic work/practice on the scale 1 – 5
- A score of 1 reflects the little or no experience
- A score of 5 reflects a high level of experience

1 2 3 4 5

Please describe how much you practice you have had in using any laparoscopic training instruments:

Eg. 30 minutes using laparoscopic training machine in Cushieri skills centre

Pre-study **trainer confidence** level:

- Please rate below your confidence in the trainer in their ability to provide feedback using the scale 1 – 5
- A score of 1 reflects the little or no confidence
- A score of 5 reflects a high level of confidence

1 2 3 4 5

Group

Appendix L: Pilot two task instruction presentation (slides)

Pilot study
Miss Sarah Gill

Feedback in Surgical Education
Task instructions

Aim of the task – Task 1

- This study features two simulated laparoscopic tasks.
- Task One is the main focus.
- The aim of this task on the LapSim is to correctly **apply two clips across a vessel** lying in the abdominal cavity, before then **dividing the vessel**.
- **Any dropped clips plus any large volumes of blood must be removed before the task is deemed completed.**

Completing the task

- Optimal task completion:
  - Upon initially grasping the vessel, two green 'clip target areas' will appear. Ideally, a clip should be placed in each of these two zones.
  - Once the clips have been applied in the green target areas, a blue 'cutting target area' will appear; the vessel should be divided in this area.

The clips can be placed outwith the green zones and will still function and the vessel can be divided anywhere between the two clips to complete the task.

How is task performance measured?

- There are several ways that task performance is measured:
  - Length of time taken to complete
  - Number of 'target areas' completed
  - Economy of instrument movement
  - Number of clips used
  - Number of clips dropped
  - Blood loss

Instruments

- You have a laparoscopic tool for each hand, left and right.
- Four types of instruments are available for you to use:
  - Grasper
  - Clip applier
  - Curved scissors
  - Suction
    - The foot pedal must be pressed to activate the suction device.
- You can change the angle of the end of the instrument by using cogwheel by your finger tips.

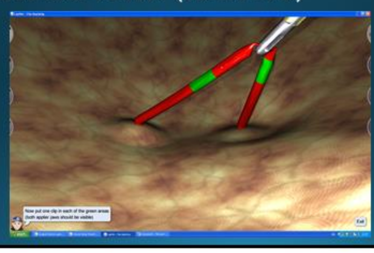
Changing instruments

- You can change the instrument used in either hand at any time:
- Pull the instrument back as far as you can (as if removing it from the abdomen)
- A menu of the four instruments will appear on the screen with the selected instrument highlighted
- Click together the handles of the instrument on that side to cycle between the instruments
- Stop when the one you want to use is highlighted, then re-insert the instrument and continue the task

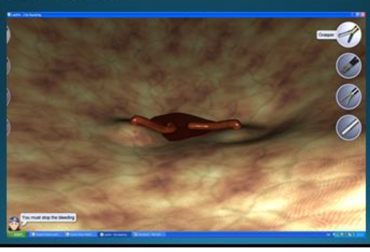
Pitfalls and complications:

- **Blood vessels are susceptible to stretch damage and can break**
- As the blood vessel is stretched, it will turn red to show it is in danger of breaking. If it is stretched beyond it's limit, it will break. The task can be completed by placing a clip on each end of the vessel to stop the bleeding.
- Even broken and clipped vessels can undergo stretch damage and the vessel wall can bleed. The bleeding is stopped by placing a clip below the level of bleeding.
- Minimise pressure on the abdominal wall from the instrument tips as this can cause damage. If the tips of the instrument are pressed into the abdominal wall, the screen will go red. This disappears once the pressure is relieved.

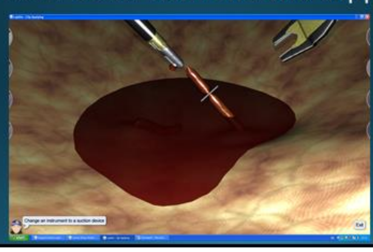
Stretched vessel (turns red)



Broken vessel



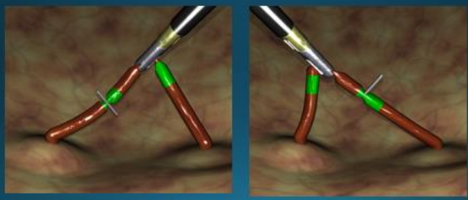
Broken vessel with one end clipped



Pitfalls and complications:

- **Clip placement**
- For the clips to be effective, it needs to be placed across the vessel with equal length of clip showing on either side.
- If the clip isn't placed correctly, it will not occlude the vessel properly and the clip won't function, meaning the vessel will bleed if divided.
- If this occurs, simply ignore this clip and place another. These clips do not have to be removed before the end of the procedure.

Functional & non-functional clips



Pitfalls and complications:

Partially closed clips

- If the jaws of the clip applier are partially closed prior to being applied to the vessel, the clip will not spring open again and it needs to be discarded by pulling the clip applier back as far as it will go, then re-insert the instrument.
- A new clip is automatically loaded for you

Dropped clips

- Clips can be accidently dropped on the abdominal floor if the jaws of the clip applier are closed when it is not applied to a vessel
- As previously stated, all dropped clips must be recovered with the graspers. Pick the clip up with the grasper and pull back as far as it will go; the clip as automatically removed from the grasper.

Dropped clips

Finishing the task

- Once the task has been completed, and all criteria have been fulfilled, the screen will simply fade out.
- **Do not press the pedal** if invited to do so **'to end the task'** as this will prematurely exit the procedure and it will not show as completed.

Further instructions

- During a 5 minute practice session, the instructor will be able to answer technical questions about the task how to operate the LapSim
- The instructor won't be able to answer questions to do with how best to approach the task!

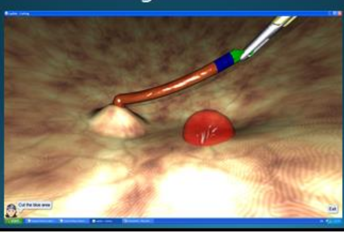
Aim of the task – Task 2

- The aim of this task on the LapSim is to cut the vessel into sections, placing the cut sections into the red target zone
- For this task, you have a grasper and a pair of cautery scissors which heat up when the foot pedal is depressed and section the vessel
- The vessel should be grasper in the green zone and cut in the blue zone. This blue zone will only appear if the vessel is stretched enough
- The cut piece of vessel is then placed in the red hemisphere, the hemisphere will turn green and the section can then be released.
- The vessel should be **sectioned 5 times** and the task will end automatically

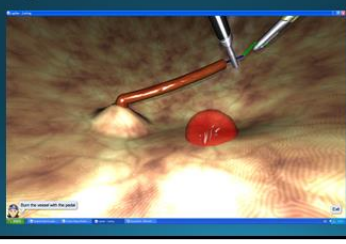
How is task performance measured?

- There are several ways that task performance is measured:
  - Length of time taken to complete
  - Number of vessel sections obtained
  - Economy of instrument movement
  - Number of times section places in hemisphere
  - The amount the vessel is stretched

The green grasping area, blue cutting zone and red target area



Cutting the vessel



Placing the cut section of the vessel into the target zone. It turns green when correctly placed.

Pitfalls and complications:

- **Vessel rupture**
- As with the previous task, the vessel will withstand a certain amount of stretch before it breaks and bleeds
- Once the vessel has ruptured, the task will terminate

Appendix M: Pilot two post study data form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Post-study data collection

Participant number:	

Post-study **confidence** level:

- Please rate below your confidence with laparoscopic tasks on the scale 1 – 5
- A score of 1 reflects the least confidence
- A score of 5 reflects the most confident

1 2 3 4 5

Post-study **experience** level:

- Please rate below your experience with laparoscopic work/practice on the scale 1 – 5
- A score of 1 reflects the little or no experience
- A score of 5 reflects a high level of experience

1 2 3 4 5

Post-study **feedback satisfaction** level:

- Please rate below your satisfaction in relation to the feedback you received as part of the study using the scale 1 – 5
- A score of 1 reflects the little satisfaction
- A score of 5 reflects a high level of satisfaction

1 2 3 4 5

Any comments:

Appendix N: Pilot two information transfer feedback form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education
Pilot study: Group A
Feedback checklist

Participant number: Task #: Level: Time for FB:

	Component	Rating			Comments
G1	Follows and agreed, logical sequence or protocol for the procedure	N	D	S	> At start > Adaptation > At end – bleeding / clip removal > Problem solving
G2	Consistently handles tissue well with minimal damage	N	D	S	> Arm position > Danger to vessel when 2 instruments being used > Inadvertent movements > Ruptured vessel
G3	Controls bleeding promptly	N	D	S	> Is 1st thought suction > Clipping held vessel before transferring to 2nd bleeder > Use of grasper for temporary arrest > Recognising bleeding vessel
G4	Uses instruments appropriately and safely	N	D	S	> Instrument selection > Dexterity > Instruments always in view > Orientation > Co-ordination of instruments
G5	Proceeds at an appropriate pace with economy of movement	N	D	S	> Bleeding - complete blood removal before changing task > Check clip good before releasing vessel > Multitasking with 2 instruments > Reducing instrument travel > 2nd hand neglect
G6	Deals calmly and effectively with unexpected events and complications	N	D	S	> Emotional reaction > Tactical reaction
T1	Grasps vessel carefully with grasper to allow clip application	N	D	S	> Position of grasper on vessel > Positioning of vessel > Holding vessel steady
T2	Applies clips correctly to either side of vessel	N	D	S	> Position of clip - Within target areas - To stop bleeding - Taught not flaccid vessel > Maximising clip efficacy / minimising BPC
T3	Cuts vessel safely with scissors	N	D	S	
T4	Retrieves dropped clips	N	D	S	> Use of 2 graspers > Use of closest grasper > Dexterity to reduce accidental clip closure

Level of performed task		
0	Unable to complete task	
1	Able to complete task with significant difficulty in several components	
2	Able to complete task with mild-moderate difficulty in several components	
3	Able to perform all elements of task with mild difficulty in only some components	
4	Able to perform all elements of task, exhibiting high skill level	

Appendix O: Pilot two dialogic feedback form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pilot study: Group B
Feedback notes

Participant number: Task #: Level: Time taken for FB:

	Component	Systematic / tactical awareness	Technical skills
G1	Logical sequencing	> At start > Adaptation > At end – bleeding / clip removal > Problem solving	
G2	Tissue handling		> Arm position > Danger to vessel when 2 instruments being used > Inadvertent movements > Ruptured vessel
G3	Use of instruments	> Instrument selection	> Dexterity > Instruments always in view > Orientation > Co-ordination of instruments
G4	Pace with economy of movement	> Bleeding - complete blood removal before changing task > Check clip good before releasing vessel	> Multitasking with 2 instruments > Reducing instrument travel > 2nd hand neglect
G5	Deals calmly and effectively with unexpected events and complications	> Emotional reaction > Tactical reaction	
G6	Awareness of self-regulation	> Tasks within tasks > Ensures clips on before releasing / cutting > Evidence of checking work	
T1	Use of grasper to allow clip application	> Position of grasper on vessel > Positioning of vessel	> Holding vessel steady, reducing
T2	Applies clips correctly	> Position of clip - Within target areas - To stop bleeding - Taught not flaccid vessel	> Maximising clip efficacy / minimising BPC > "Sneaking up"
T3	Control of bleeding	> Is 1st thought suction > Clipping held vessel before transferring to 2nd bleeder > Use of grasper for temporary arrest	> Recognising bleeding vessel
T4	Minimises dropped clips / Optimises method of retrieval	> Use of 2 graspers > Use of closest grasper > Awareness of number dropped	> Dexterity to reduce accidental clip closure

GOALS: **1.**

2.

Appendix P: Statistical calculations

MD Feedback in Surgical Education: Final study statistical Power Calculations

Summary:

- Calculations based upon experimental data relating to P4 (pilot study two)
- Data input in red; generated values are in black
- Calculated sample sizes are **bolded**

Time Taken to Complete task

Confidence Interval (2-sided)	95%
Power	80%
Ratio of sample size (Group 2/Group 1)	1:1

- **Sample Size for Comparing Two Means: Experimental data only**

	Group 1	Group 2	Difference*
Mean	270	100	170
Standard deviation	202	18	
Variance	40804	324	

Sample size of Group 1	**13**
Sample size of Group 2	**13**
Total sample size	**26**

- **Sample Size for Comparing Two Means: Experimental data plus minimally significant difference set to 60 seconds**

	Group 1	Group 2	Difference*
Mean			60
Standard deviation	202	18	
Variance	40804	324	

Sample size of Group 1	**90**
Sample size of Group 2	**90**
Total sample size	**180**

- **Sample Size for Comparing Two Means: Experimental data plus minimally significant difference set to 120 seconds**

	Group 1	Group 2	Difference*
Mean			120
Standard deviation	202	18	
Variance	40804	324	

Sample size of Group 1	**23**
Sample size of Group 2	**23**
Total sample size	**46**

Combined Instrument Path Length

Confidence Interval (2-sided)	95%
Power	80%
Ratio of sample size (Group 2/Group 1)	1:1

- **Sample Size for Comparing Two Means: Experimental data only**

	Group 1	Group 2	Difference*
Mean	6.77	2.41	4.36
Standard deviation	4.88	0.94	
Variance	23.8144	0.8836	

Sample size of Group 1	**11**
Sample size of Group 2	**11**
Total sample size	**22**

- **Sample Size for Comparing Two Means: Experimental data plus minimally significant difference set to 1 metre**

	Group 1	Group 2	Difference*
Mean			1
Standard deviation	4.88	0.94	
Variance	23.8144	0.8836	

Sample size of Group 1	**194**
Sample size of Group 2	**194**
Total sample size	**388**

- **Sample Size for Comparing Two Means: Experimental data plus minimally significant difference set to 2.5 metres**

	Group 1	Group 2	Difference*
Mean			3
Standard deviation	4.88	0.94	
Variance	23.8144	0.8836	

Sample size of Group 1	**22**
Sample size of Group 2	**22**
Total sample size	**44**

<u>Number of Bad Clips Placed</u>

Confidence Interval (2-sided)	95%
Power	80%
Ratio of sample size (Group 2/Group 1)	1:1

- **Sample Size for Comparing Two Means: Experimental data only**

	Group 1	Group 2	Difference*
Mean	5.67	2	3.67
Standard deviation	5.96	1.1	
Variance	35.5216	1.21	

Sample size of Group 1	**22**
Sample size of Group 2	**22**
Total sample size	**44**

- **Sample Size for Comparing Two Means: Experimental data plus minimally significant difference set to 2 BPC**

	Group 1	Group 2	Difference*
Mean			2
Standard deviation	5.96	1.1	
Variance	35.5216	1.21	

Sample size of Group 1	**73**
Sample size of Group 2	**73**
Total sample size	**146**

- **Sample Size for Comparing Two Means: Experimental data plus minimally significant difference set to 3 BPC**

	Group 1	Group 2	Difference*
Mean			3
Standard deviation	5.96	1.1	
Variance	35.5216	1.21	

Sample size of Group 1	**33**
Sample size of Group 2	**33**
Total sample size	**66**

Appendix Q: Full study participant information sheet

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

RESEARCH STUDY PARTICIPANT INFORMATION SHEET

INVITATION TO TAKE PART IN A RESEARCH STUDY
You are being invited to take part in a research study, which aims to investigate the possible relationship between feedback and surgical skill development. The lead investigator is Miss Sarah Gill, a surgical trainee in the East of Scotland Deanery, who is undertaking this study as part of an MD in Medical Education. The study is being supervised by Drs Rola Ajjawi and Sean McAleer from the Centre for Medical Education and Mr Fraser Harold Orthopaedic Surgeon at Ninewells Hospital.

WHAT IT INVOLVES & THE PURPOSE OF THE RESEARCH STUDY
Participation in this study will involve performing a surgical task on a laparoscopic simulator and participation in feedback sessions in relation to task performance. This process will be repeated over several sessions. The task will be explained prior to you undertaking it and your performance will be measured but this will remain confidential between only you and the investigator. One of these performances will be recorded and reviewed by yourself and the lead researcher. During this session, you will be interviewed about the task and this interview will be recorded. All data will be stored anonymously. The relationship between feedback and performance is what is of interest. It is not possible to "fail" in the study as it is not a test.

It is proposed that investigation into this field and your participation in this research would enhance our understanding of feedback in surgical education and so benefit future surgical trainees. We hope that the participants in this study will find the experience interesting (and fun) and will benefit from the surgical skills acquired as part of the study.

TIME COMMITMENT
The study will require four sessions in total. The first three sessions will last approximately 35-50 minutes and the final visit will last approximately 15 minutes. The first three visits will be one week apart and the final visit four weeks after the third visit. These will be arranged in advance at a mutually convenient time for both the participant and the investigator. The study will take place in a test room in the Tayside Orthopaedic & Rehabilitation Technology Centre at Ninewells Hospital.

COST, REIMBURSEMENT AND COMPENSATION
Your participation in this study is voluntary and of course participating in the study does not cost anything. You will be awarded a certificate of participation for your portfolio at the fourth session as thanks for your commitment to the study.

RISKS AND TERMINATION OF PARTICIPATION
There are no known risks for you in this study. You may decide to stop being a part of the research study at any time without explanation and without penalty.

CONFIDENTIALITY/ANONYMITY
The data collected will include demographics, such as age and gender, and experimental results but no one will be able to link this data to your identity and name. The data will be seen only by the researchers and will not be made available to anyone else. The anonymised results will be published as part of an MD thesis and research papers.

FOR FURTHER INFORMATION ABOUT THIS RESEARCH STUDY
Sarah Gill will be glad to answer your questions about this study at any time and can be contacted using the details below. Should you wish to find out about the final results of this study, she will be happy to take your contact details and send you a copy of the results.

Address: Medical Education Directorate, Level 8, Dundee Medical School, Ninewells Hospital, Dundee, DD1 9SY
Email: s.v.gill@dundee.ac.uk
Telephone: 01382 383528

The University Research Ethics Committee of the University of Dundee has reviewed and approved this research study.

Appendix R: Full study participant consent form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

RESEARCH STUDY PARTICIPANT CONSENT FORM

This research study aims to investigate the possible relationship between feedback and surgical skill development. Participation in this study will involve performing a surgical task on a laparoscopic simulator and participation in a feedback session in relation to performance. This process will be repeated over several sessions.

By signing below you are indicating that you have read and understood the Participant Information Sheet and that you agree to take part in this research study.

I agree to the use of anonymous extracts from my interview in conference papers and academic publications	YES	☐	NO	☐
I agree to the video recording of task performed on the laparoscopic simulator	YES	☐	NO	☐
I agree to the audio recording of interviews	YES	☐	NO	☐

Signed (participant): .... **Date:** ....

Name (participant): ....

Signed (investigator): .... **Date:** ....

Name (investigator): Miss Sarah Gill

Appendix S: Full study pre-study data form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Pre-study data collection

Date:	
Participant number:	
Sex:	
Age:	
Hand dominance:	

Pre-study **confidence**:

- Please rate below your **confidence** with **laparoscopic tasks in general** on the scale 1 – 5

1 2 3 4 5

Not at all confident — Extremely confident

Pre-study **experience**:

- Please rate below your **level of experience** with **laparoscopic work/practice in general** on the scale 1 – 5

1 2 3 4 5

No experience at all — Extensive experience

Please describe any previous experience with laparoscopy, including practice using laparoscopic training instruments:

Eg. 30 minutes using laparoscopic training machine in Cushieri skills centre

Setting Theatre, Cushieri skills centre, surgical course, surgical society event	**Instruments used** Real laparoscopic instruments Laparoscopic model with live camera Laparoscopic simulator (PC screen)	**Duration** Number of minutes for each session x number of sessions
Eg. Cushieri skills centre	*Laparoscopic simulator*	*30 mins*

This section of the form is kept separate to the other pre-study data. It is not linked to your participation number, only to the larger group you are part of.

Once you have completed it, the instructor will show you the box to place it in. The box will not be opened until the end of the study and all of the results are anonymised.

Pre-study **confidence in trainer** level:

- Please rate below your **confidence in the trainer** in their ability to provide feedback, using the scale 1 – 5

	1	2	3	4	5	
Not at all confident						Extremely confident

Group

A / B

Appendix T: Full study post study data form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Post-study data collection

Participant number:	

Post-study **confidence**:

- Please rate below your **confidence** with **laparoscopic tasks in general** on the scale 1 – 5

1 2 3 4 5

Not at all confident Extremely confident

Post-study **confidence in tutor**:

- Please rate below your **confidence in the tutor** in their ability to provide feedback, using the scale 1 – 5

1 2 3 4 5

Not at all confident Extremely confident

Post-study **satisfaction**:

- Please rate below your **level of satisfaction** relating to participating in this study using the scale 1 – 5

1 2 3 4 5

Not at all satisfied Extremely satisfied

Please tell us why:

Would you recommend participating in this study to a friend? (Please circle)

YES *NO*

- In relation to the feedback sessions, please rate below how well you feel the forms used helped to provide you with useful feedback

1 2 3 4 5

Not at all helpful Extremely helpful

Appendix U: Full study information transfer feedback form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education
Pilot study: Group A
Feedback Type A checklist

Participant number: Performance #: Time for FB:

	Component	Rating			Comments
G1	Follows and agreed, logical sequence or protocol for the procedure	N	D	S	> At start > Adaptation > At end – bleeding / clip removal > Problem solving
G2	Consistently handles tissue well with minimal damage	N	D	S	> Arm position > Danger to vessel when 2 instruments being used > Inadvertent movements > Ruptured vessel
G3	Controls bleeding promptly	N	D	S	> Is 1st thought suction > Clipping held vessel before transferring to 2nd bleeder > Use of grasper for temporary arrest > Recognising bleeding vessel
G4	Uses instruments appropriately and safely	N	D	S	> Instrument selection (Early) > Dexterity > Instruments always in view > Orientation > Co-ordination of instruments
G5	Proceeds at an appropriate pace with economy of movement	N	D	S	> Bleeding - complete blood removal before changing task > Check clip good before releasing vessel > Multitasking with 2 instruments > Reducing instrument travel > 2nd hand neglect
G6	Deals calmly and effectively with unexpected events and complications	N	D	S	> Emotional reaction > Tactical reaction
T1	Grasps vessel carefully with grasper to allow clip application	N	D	S	> Position of grasper on vessel > Positioning of vessel > Holding vessel steady
T2	Applies clips correctly to either side of vessel	N	D	S	> Position of clip - Within target areas - To stop bleeding - Taught not flaccid vessel end > Optimising clip technique - Adequate tension - Vessel properly in jaws - Seeing both jaws of clipper - Orientation of clipper
T3	Cuts vessel safely with scissors	N	D	S	
T4	Retrieves dropped clips	N	D	S	> Use of 2 graspers > Use of closest grasper > Dexterity to reduce accidental clip closure

Level of performed task		
0	Unable to complete task	
1	Able to complete task with significant difficulty in several components	
2	Able to complete task with mild-moderate difficulty in several components	
3	Able to perform all elements of task with mild difficulty in only some components	
4	Able to perform all elements of task, exhibiting high skill level	

Appendix V: Full study dialogic feedback form

THE UNIVERSITY OF DUNDEE: DEPARTMENT OF MEDICAL EDUCATION
Feedback and learning in surgical education

Feedback Type B Form / notes

Participant number: Performance #: Time taken for FB:

	Component	Systematic / tactical awareness	Technical skills
G1	Logical sequencing	> At start > Adaptation > At end – bleeding / clip removal > Problem solving	
G2	Tissue handling		> Arm position > Danger to vessel when 2 instruments being used > Inadvertent movements > Ruptured vessel
G3	Use of instruments	> Instrument selection (Early)	> Dexterity > Instruments always in view > Orientation > Co-ordination of instruments
G4	Pace with economy of movement	> Bleeding - complete blood removal before changing task > Check clip good before releasing vessel	> Multitasking with 2 instruments > Reducing instrument travel > 2nd hand neglect
G5	Deals calmly and effectively with unexpected events and complications	> Emotional reaction > Tactical reaction	
G6	Awareness of self-regulation	> Were you monitoring your performance? > How were you monitoring? What were you checking for? > Was there evidence of your goals/aims in this performance? > Evidence of checking work	
T1	Use of grasper to allow clip application	> Position of grasper on vessel > Positioning of vessel	> Holding vessel steady, reducing accidental movement
T2	Applies clips correctly	> Position of clip - Within target areas - To stop bleeding - Taught not flaccid vessel end	> Optimising clip technique - Adequate tension - Vessel properly in jaws - Seeing both jaws of clipper - Orientation of clipper > "Sneaking up"
T3	Control of bleeding	> Is 1st thought suction > Clipping held vessel before transferring to 2nd bleeder > Use of grasper for temporary arrest	> Recognising bleeding vessel
T4	Minimises dropped clips / Optimises method of retrieval	> Use of 2 graspers > Use of closest grasper > Awareness of number dropped	> Dexterity to reduce accidental clip closure

GOALS/AIMS: 1.

2.